

An Introduction to the Languages of the World

Second Edition



Anatole V. Lyovin • Brett Kessler • William R. Leben

An Introduction to the Languages of the World

An Introduction to the Languages of the World

SECOND EDITION

Anatole V. Lyovin
Brett Kessler
and
William R. Leben

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

© Oxford University Press 2017

First Edition published in 1997
Second Edition published in 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data

Names: Lyovin, Anatole, author. | Kessler, Brett, 1956– author. |
Leben, William Ronald, 1943– author.

Title: An introduction to the languages of the world / Anatole V. Lyovin,
Brett Kessler, William R. Leben.

Description: Oxford ; New York : Oxford University Press, [2017] |
Includes bibliographical references and index.

Identifiers: LCCN 2016033660 (print) | LCCN 2016041209 (ebook) |
ISBN 9780195149883 (pbk. : alk. paper) | ISBN 9780190645267 (pdf) |
ISBN 9780190645274 (ebook) | ISBN 9780190662752 (online resource)

Subjects: LCSH: Language and languages. | Linguistics.

Classification: LCC P371 .L96 2017 (print) | LCC P371 (ebook) | DDC 410—dc23
LC record available at <https://lccn.loc.gov/20160336>

9 8 7 6 5 4 3 2 1

Printed by Sheridan Books, Inc., United States of America

PREFACE

Preface to the first edition

This book is intended to be used in a course designed for students who have mastered the basic principles of linguistics but lack background information about the broad range of language phenomena (vowel harmony, ergative constructions, etc.) found in the world's languages, and who also need to learn a few facts about the existence of major language families, the distribution of major language groups, and so forth. It aims to offer students an opportunity to explore, at various levels, structures of very different, highly interesting languages without necessarily possessing a speaking or reading knowledge of these languages, as they would in normal language classes. At the same time, this book can serve as an introduction to language typology in general.

It should be emphasized that this text is not intended as an introduction to linguistics, since it presupposes some familiarity with such basic linguistics concepts as morpheme, phoneme, and basic articulatory phonetics terms. It is intended as a textbook for undergraduates who have already taken at least one advanced course in general linguistics and may also have taken an introductory course in phonetics. It is also intended for graduate students in linguistics who have already taken their required introductory courses in various branches of linguistics but otherwise have little background knowledge of the field. At the University of Hawaii at Mānoa, for example, a course for which this work would be assigned as a textbook requires as a minimum prerequisite an upper-division introductory course in general linguistics, which beginning graduate students who have no background in linguistics are required to take. Naturally, students who have had other linguistics courses (e.g. phonetics, phonology, and syntax) usually enjoy this course more and encounter fewer difficulties.

Even though this book is not intended as an introductory text in linguistics, certain concepts are reviewed in some detail in the introductory chapter because my experience as a teacher of beginning linguistics students has taught me to expect all sorts of misconceptions about what linguists mean when they say that certain languages are "genetically related". It is also not intended as a reference book, since it is limited in scope and coverage and certainly cannot compete with such books as Bernard Comrie's *The World's Major Languages* (New York: Oxford University Press, 1987), which contains in-depth articles on major languages and language families written by recognized experts in their respective fields. Lacking expertise in all the languages of the world, I have had to depend on the opinions of others

without always being able to check the validity and accuracy of the latter. On the other hand, reference works are usually too technical and contain too many details for beginning students. In addition, they are not ideal as textbooks because, among other things, they do not contain such things as exercises and other addenda helpful to the instructor as well as students in introductory-level classes.

The need for such a textbook became apparent about twenty years ago, when I began teaching a course on the languages of the world and found that there were simply no suitable textbooks for such courses, despite the existence of several books entitled *Languages of the World*. (Most such books deal only with different writing systems found throughout the world.) It is only in recent years that up-to-date reference books on the languages of the world have appeared that are suitable to be placed on the reserve shelf for courses on the subject.

This book surveys all the areas of the globe on a continent-by-continent basis. It does not concentrate on major languages, although most of the languages in the sketches may be considered so. Some languages were chosen because of their typological interest and not the socioeconomic prominence of their speakers. However, not all language types are represented. This would have required that the scope of this textbook be greatly expanded; as it is, it would be too difficult to cover the entire contents adequately in one semester. It would probably be best for the instructor to choose, on the basis of the interests of the students, which language sketches to assign and which to skip. It is hoped that students will read independently any materials not formally covered during the course.

The coverage is not intended to be encyclopedic and exhaustive because there exist many excellent reference sources that may be consulted by students wishing to know more details about classification of particular languages, geographical distribution of languages, and the number of speakers of individual languages. My aim has been to strike a balance between overwhelming the readers with too much detailed information about various languages and the controversies involving their genetic classification, and giving too superficial an account. To make the account simpler and shorter, the coverage emphasizes the living languages rather than the extinct ones, although on occasion references are made to extinct languages as well. At the end of each chapter, there is a bibliography which will provide a good basis for readers who want to know more about some specific topic.

As for the controversial topics that have stirred the linguistic community in the recent and not-so-recent past, I felt that it would be highly remiss to avoid any mention of Joseph Greenberg's Amerind and Indo-Pacific hypotheses, Paul Benedict's Austro-Tai and Austro-Tai-Japanese hypotheses, and the Nostratic hypothesis being championed by Vitaly Shevoroshkin and colleagues. In the case of the Altaic hypothesis, which already has a venerable history and cannot be considered a recent development, it seemed a disservice to the readers to dismiss the whole matter with a mere mention in a footnote. Thus, all of these ideas are mentioned and briefly described in this text, and bibliographical references are given for scholarly articles that argue both for and against these hypotheses. On

the other hand, since this is, after all, an introductory textbook, it would be out of place to include lengthy accounts of such controversies.

Some exercises may prove too difficult for students with very little background in linguistics. Such exercises may be skipped entirely, or the instructor may provide clues and hints to the students in order to help them solve the problems more easily.

I would like to express my deep gratitude to the following individuals, who helped me in the preparation of this book: Riika Alanen, who helped me with the sample Finnish text; Greg Lee Carter, who suggested many key improvements for the sketch of Hawaiian; Emily Hawkins, who helped me in the selection of the sample Hawaiian text; Darius Kenyi Jonathan, who helped with the sketch of Arabic; Gillian Sankoff, who kindly provided the text for the Tok Pisin sketch and explained what it meant; Craig Volker, who kindly reviewed my sketch of Tok Pisin; Robert Blust, who clarified for me his classification of Austronesian languages; and Michael Forman, who offered many useful suggestions, especially on the chapter dealing with pidgin and creole languages. Thanks also go to Marion Sonomura. Finally, I must acknowledge the moral support provided by Byron W. Bender, my department chair at the University of Hawaii at Mānoa, in particular his unflagging faith that I would successfully complete this project. The anonymous readers for Oxford University Press critically reviewed an earlier draft of this book and provided many constructive comments. To all these kind people I say a heartfelt “Mahalo.” If there are errors and omissions in this book, they are due to my lack of expertise in many areas.

Honolulu
January 1996
A.V.L.

Preface to the second edition

Authors Leben and Kessler are grateful to Oxford University Press for inviting us to prepare this thorough revision of the first edition of Anatole Lyovin’s *An Introduction to the Languages of the World*.

The information in the first edition has been completely updated to what is known in 2016. This includes the names of languages, the number of speakers, and their locations. All maps have been redone. They now locate all the languages discussed in the book and use the same classification scheme used in the book. The introduction to pidgins and creoles in Chapter 8 is joined by new sections on sign languages and on language attrition and revitalization.

Throughout, language names now appear with their international standard (ISO) three-letter code. This code uniquely identifies each language, making it easier to look up a language online in sources such as *Ethnologue.com* and *Glottolog.org*. Except where we note otherwise, we have adopted *Ethnologue*.

com as our primary authority for each language's name and number of speakers. Where possible, we have moved this type of information out of the text and into tables. Instead of exact estimates of numbers of speakers, which change very quickly and are hard to retain in the reader's memory, we break down population sizes into ten categories. Rather than list all regions where a language is spoken, we give the one area where the most speakers live. We recommend consulting trusted sources – including *Ethnologue.com* – for additional details and updates.

We use a standard alphabet, the International Phonetic Alphabet, to transcribe examples from all the languages in the text. We also have made an effort to standardize linguistic terminology across languages. A consistent transcription and terminology should make for easier learning and more clarity and should help bring out the real similarities and differences among languages.

All linguistic examples and texts are now aligned to show which gloss goes with which word. These examples now follow standard linguistic format and use standard linguistic abbreviations. Throughout, we have adhered to the *Generic Style Rules for Linguistics*, Martin Haspelmath's harmonization of existing style guidelines.

At the end of each chapter, the suggested readings have been updated and annotated, and several exercises have been added or revised. As in the first edition, we also include a bibliography at the end of the book, which gives complete references for the works referred to in the text.

We hope that all of these changes will help to make this introduction to the world's languages more straightforward to follow, easier to learn and remember, and more satisfying to read.

We would like to express our thanks to Daniel Leben-Wolf for Figure 2.1 and other images in Chapter 2; to Leanne Hinton for reviewing the section on North America in Chapter 7 and to Lev Michael for reviewing the section on South America; to James A. Matisoff for reviewing §4.3 on Sino-Tibetan; and to Lev Blumenfeld for reviewing the Sketch of Russian §3.5. Our illustrations were made with the help of many free and open software and data projects, whose contributions we gratefully acknowledge. These include ImageMagick, Inkscape, Graphviz, and Mapnik software, Natural Earth map data, and DejaVu and Linux Libertine fonts.

CONTENTS

List of Tables	xiii
List of Figures	xvii
Symbols and Abbreviations	xix
General symbols	xix
General abbreviations	xxi
Glossing abbreviations	xxi
Language codes	xxii
Phonetic symbols	xxxv
1 Classification of Languages	1
1.1 Genetic classification	1
1.2 Typological classification of languages	15
1.3 Exercises	26
1.4 Suggested readings	28
2 Classification of Writing Systems	31
2.1 Typological classification of writing systems	31
2.2 Genetic classification of writing systems	44
2.3 Exercises	51
2.4 Suggested readings	53
3 Europe	55
3.1 Indo-European	56
3.2 Uralic	71
3.3 Caucasus area	75
3.4 Basque	79
3.5 Sketch of Russian	80
3.6 Sketch of Finnish	103
3.7 Exercises	121
3.8 Suggested readings	125
4 Asia	127
4.1 Altaic area	127
4.2 Paleosiberian area	136
4.3 Sino-Tibetan	137
4.4 Hmong-Mien	142
4.5 Tai-Kadai	144

4.6	Austroasiatic	145
4.7	Dravidian	147
4.8	Burushaski	148
4.9	Other languages in Asia	149
4.10	Sketch of Mandarin Chinese	149
4.11	Sketch of Classical Tibetan	165
4.12	Exercises	188
4.13	Suggested readings	194
5	Africa	197
5.1	Afro-Asiatic	198
5.2	Nilo-Saharan	208
5.3	Niger-Congo	212
5.4	Khoisan area	220
5.5	Other languages in Africa	222
5.6	Sketch of Modern Standard Arabic	223
5.7	Sketch of Swahili	241
5.8	Exercises	259
5.9	Suggested readings	263
6	Oceania	265
6.1	Austronesian	265
6.2	Papuan area	272
6.3	Australian area	274
6.4	Sketch of Hawaiian	278
6.5	Sketch of Dyrirbal	298
6.6	Exercises	318
6.7	Suggested readings	322
7	The Americas	325
	North American area	325
7.1	Eskimo-Aleut	327
7.2	Na-Dene	330
7.3	Algic	332
7.4	Muskogean	335
7.5	Siouan	336
7.6	Iroquoian	337
7.7	Caddoan	338
7.8	Yuman	338
7.9	Sahaptian	339
7.10	Tsimshianic	340
7.11	Kiowa-Tanoan	340
7.12	Uto-Aztecan	341
7.13	Salishan	343

Mesoamerican area 344

7.14 Oto-Manguean 346

7.15 Totonacan 346

7.16 Mixe-Zoquean 346

7.17 Mayan 349

South American area 349

7.18 Intermediate area 351

7.19 Western Amazonia 352

7.20 Northern foothills 353

7.21 Andes region 356

7.22 Southern foothills 357

7.23 South 357

7.24 Central Amazonia 360

7.25 Northern Amazonia 362

7.26 Sketch of Central Alaskan Yup'ik 363

7.27 Sketch of Ayacucho Quechua 372

7.28 Exercises 394

7.29 Suggested readings 397

8 Language Birth, Death, and Revitalization 401

8.1 Mixed languages 401

8.2 Constructed languages 403

8.3 Pidgins 404

8.4 Creoles 409

8.5 Sign languages 415

8.6 Language endangerment and renewal 423

8.7 Sketch of Tok Pisin 427

8.8 Exercises 445

8.9 Suggested readings 447

References 451

Language Index 467

Subject Index 483

LIST OF TABLES

- 1.1 Vocabulary of four languages of the Pacific 12
- 1.2 Recurrent correspondences from Table 1.1 12
- 1.3 Data for Exercise 1 27

- 2.1 Types of graphic signs 32
- 2.2 Scripts of the world 47
- 2.3 Canaanite family of scripts 50

- 3.1 Languages of Europe 56
- 3.2 Indo-European languages 56
- 3.3 Celtic languages 58
- 3.4 Germanic languages 60
- 3.5 Romance languages 64
- 3.6 Balto-Slavic languages 65
- 3.7 Indo-Iranian languages 67
- 3.8 Development of */k/ in Indo-European languages 70
- 3.9 Basic correspondences of voiceless velar stops in the branches of Indo-European 70
- 3.10 Uralic languages 72
- 3.11 Languages of the Caucasus 75
- 3.12 Consonant phonemes of Russian 81
- 3.13 Vowel phonemes of Russian 83
- 3.14 Noun paradigms 87
- 3.15 Sample adjective declension 88
- 3.16 Personal verb forms in Russian 91
- 3.17 Consonant phonemes of Finnish 104
- 3.18 Vowel phonemes of Finnish 104
- 3.19 Consonant gradation in Finnish nouns 106
- 3.20 Finnish cases 108
- 3.21 Personal endings on Finnish verbs 112
- 3.22 Imperative mood in Finnish 113

- 4.1 Languages of Asia 127
- 4.2 Altaic languages 130
- 4.3 Paleosiberian languages 136
- 4.4 Sinitic languages 139
- 4.5 Tibeto-Burman languages 141
- 4.6 Hmong-Mien languages 143
- 4.7 Tai-Kadai languages 144

- 4.8 Austroasiatic languages 146
- 4.9 Dravidian languages 147
- 4.10 Mandarin Chinese consonant phonemes 152
- 4.11 Pronunciation of Chinese rimes 153
- 4.12 The tones of Mandarin Chinese 155
- 4.13 Personal pronouns of Mandarin Chinese 156
- 4.14 Tibetan languages 167
- 4.15 Tibetan grammatical relation markers 173
- 4.16 Honorifics in Tibetan 181
- 4.17 Old Tibetan and Modern Lhasa Tibetan 189
- 4.18 Hmong and Mandarin Chinese 190

- 5.1 Languages of Africa 197
- 5.2 Afro-Asiatic languages 200
- 5.3 Semitic languages 200
- 5.4 Berber languages 203
- 5.5 Chadic languages 204
- 5.6 Cushitic, Omotic, and Egyptian languages 205
- 5.7 Nilo-Saharan languages 209
- 5.8 Northwestern Niger-Congo languages 214
- 5.9 Benue-Congo languages 216
- 5.10 Khoisan languages 220
- 5.11 Symbols used to transcribe clicks 221
- 5.12 Consonant phonemes of Modern Standard Arabic 225
- 5.13 Changes in letter shapes between Phoenician and Arabic 226
- 5.14 The Arabic alphabet 227
- 5.15 Arabic noun inflections 230
- 5.16 Arabic personal pronouns, free forms 233
- 5.17 Arabic personal enclitic pronouns 233
- 5.18 Verb derivation patterns in Arabic 234
- 5.19 Arabic active indicative verb inflections 235
- 5.20 Consonants of Swahili 242
- 5.21 Swahili genders 244
- 5.22 Swahili gender markers on nouns 245

- 6.1 Austronesian languages 267
- 6.2 Papuan languages 273
- 6.3 Australian languages 275
- 6.4 Hawaiian consonant phonemes 279
- 6.5 Vowel phonemes of Hawaiian 280
- 6.6 Hawaiian personal pronouns 283
- 6.7 Case proclitics in Hawaiian 286
- 6.8 Consonant phonemes of Dyirbal 300
- 6.9 Dyirbal case inflections 302
- 6.10 Dyirbal personal pronouns 306

- 6.11 Dyirbal verb inflections 306
- 6.12 Dyirbal verb suffixes 307
- 6.13 Buang data for Exercise 4 321
- 6.14 Kiwai data for Exercise 5 321

- 7.1 Languages of North America 326
- 7.2 Eskimo-Aleut languages 330
- 7.3 Na-Dene languages 331
- 7.4 Algic languages 333
- 7.5 Muskogean languages 335
- 7.6 Siouan languages 336
- 7.7 Iroquoian languages 337
- 7.8 Caddoan languages 338
- 7.9 Yuman languages 338
- 7.10 Sahaptian languages 339
- 7.11 Tsimshianic languages 340
- 7.12 Kiowa-Tanoan languages 341
- 7.13 Uto-Aztecan languages 342
- 7.14 Salishan languages 343
- 7.15 Languages of Mesoamerica 345
- 7.16 Oto-Manguean languages 346
- 7.17 Totonacan languages 348
- 7.18 Mixe-Zoquean languages 348
- 7.19 Mayan languages 350
- 7.20 Languages of South America 351
- 7.21 Languages of the Intermediate Area 354
- 7.22 Languages of Western Amazonia 354
- 7.23 Languages of the northern foothills of the Andes 356
- 7.24 Languages of the Andes 357
- 7.25 Languages of the southern foothills of the Andes 358
- 7.26 Languages of southern South America 359
- 7.27 Languages of Central Amazonia 360
- 7.28 Languages of Northern Amazonia 363
- 7.29 Yup'ik consonants 365
- 7.30 Yup'ik case endings 366
- 7.31 Quechua consonants 374
- 7.32 Case endings in Ayacucho Quechua 376
- 7.33 Quechua personal pronouns 379

- 8.1 Mixed languages 403
- 8.2 Pidgins 409
- 8.3 Creoles 413
- 8.4 Sign languages 416
- 8.5 Tok Pisin consonants 429
- 8.6 Tok Pisin personal pronouns 430

LIST OF FIGURES

- 0.1 IPA chart xxxiv
- 1.1 Anagenesis 2
- 1.2 Cladogenesis 5
- 1.3 Cladistic arrangement of four Pacific languages 14
- 2.1 Iconic representation 33
- 2.2 Graphic symbols 33
- 2.3 AIGA DOT semasiograms 52
- 3.1 Indo-European languages 57
- 3.2 Uralic languages 72
- 3.3 Languages of the Caucasus 76
- 4.1 Languages of North Asia 128
- 4.2 Languages of South Asia 138
- 4.3 Mandarin Chinese syllable structure 152
- 4.4 Stamps depicting The Foolish Man from Chu 164
- 5.1 Languages of northern Africa 199
- 5.2 Languages of southern Africa 213
- 6.1 Languages of Oceania 266
- 7.1 Languages of North America 328
- 7.2 Languages of Mesoamerica 345
- 7.3 Languages of South America 352

SYMBOLS AND ABBREVIATIONS

General symbols

We follow standard linguistic symbols and abbreviations whenever they are available. In many cases, the *Leipzig Glossing Rules* (Max Planck Institute 2008) explain these conventions with additional examples.

italics This font indicates that content is being mentioned, cited, or referred to: “The first-person singular of *is* is *am*.” Italics are not used in tables or displayed examples if the context makes it clear that a word is being cited.

SMALL CAPS This font is used when a technical term is introduced: “Words for sentient beings have ANIMATE gender.” It is also used for abbreviations in grammatical analyses, such as AN, the abbreviation for the animate gender. We also use this font for the three-letter ISO symbols for languages: the code for Choctaw is CHO.

[] Square brackets enclose pronunciations in IPA (defined below): “English *pig* is pronounced [p^hi:ɡ].” Elements in brackets represent PHONES, the individual segments of speech. When discussing language structure in this book, we mostly transcribe phones at the level of PHONEMES: we only use distinct symbols for sounds that are lexically contrastive in a particular language.

In a grammatical analysis, square brackets enclose a category that is not explicitly indicated by a separate morpheme: Latin *puer* can be analyzed with the abbreviation for *singular* in square brackets because nouns in Latin are singular when they do not have a plural ending.

puer
boy[sg]

() In a grammatical analysis, parentheses surround a category that is not indicated by a separate morpheme, but is an inherent part of the preceding morpheme: Latin *mater* can be analyzed with the abbreviation for *feminine* in parentheses, because that word is always feminine.

ma:ter
mother(F)

In other contexts, numbers in parentheses refer to displayed examples. Numbering of examples starts with (1) in each chapter. References refer to examples in the same chapter unless otherwise stated.

- ⟨ ⟩ In a grammatical analysis, angled quotes denote INFIXES: morphemes that are inserted inside other morphemes. In Choctaw, [n] is inserted into verbs like [talaja] to show that a state continues (durative aspect).

talá⟨n⟩ja
sit⟨DUR⟩

A single > means that a single morpheme marks both subject agreement (shown to the left of the >) and object agreement (shown to the right).

In other contexts, angled quotes enclose content that is being cited as characters of writing: “The doubled ⟨aa⟩ in ⟨romaani⟩ ‘novel’ indicates that the vowel sound is long.”

- ‘ ’ Single quotes enclose meanings: “The Irish word *bó* ‘cow’ is irregular.” Translations of examples are also enclosed in single quotes.

≡ means ‘corresponds to’.

- separates morphemes: “*un-break-able*”. Prefixes cited alone have a hyphen after them, while suffixes have a hyphen before them: “The word *unbreakable* has the prefix *un-* and the suffix *-able*.”

- ~ separates a reduplicant from its base. A reduplicant is an affix whose phonemes are wholly or partly copied from the base: “The preterite of Latin *curr-o* ‘I run’ is *cu~curr-i*.”

- = separates a clitic from its base. A clitic is a morpheme that functions syntactically like a grammatical word but phonologically more like an affix. For example, in Latin [=kwɛ] can attach to any word to mean ‘and’:

arma wirum=kwɛ kano:
weapons man=and sing.1SG
‘I sing of weapons and a man.’

- \ In a grammatical analysis, the backslash separates a morpheme from its base when the affix is some alteration of the base. English *men* can be analyzed with backslash and the abbreviation for *plural* because the plural number is indicated by changing the vowel of the base from *a* to *e*.

mɛn
man\PL

- . In a morphological analysis, a dot separates elements that together are expressed by one morpheme in the language being described. This analysis of *-us* uses dots to show that the suffix simultaneously marks masculine gender, singular number, and nominative case.

serw-us
slave-M.SG.NOM

∅ indicates the absence of a phone or morpheme in a context where other words or other languages have a phone or morpheme.

1, 2, 3, etc. When numbers are used to indicate the size of a speaker population, they count the number of digits needed to represent the number of speakers. Thus 1 means 1 to 9 speakers, 2 means 10 to 99, and so forth.

* means that what follows would be ungrammatical: “The past tense of *go* is *went*, not **goed*.”

* appears to the left of a form that has been reconstructed from the past, when there is no textual record of it.

§ refers to a numbered section of this book.

General abbreviations

IPA	International Phonetic Alphabet
ISO	International Organization for Standardization
O	object
S	subject
V	verb

Glossing abbreviations

1	first person	ADV	adverb	AUG	augmentative
2	second person	AGGR	aggregate	AUX	auxiliary
3	third person	AGT	agent	BEN	benefactive
ABE	abessive	ALIEN	alienable	CAUS	causative
ABL	ablative	ALL	allative	CAUSE	causal
ABS	absolutive	AN	animate	CLF	classifier
ABST	abstract	AND	andative	COM	comitative
ACC	accusative	ANTIP	antipassive	COMP	comparative
ACT	active	APPL	applicative	COND	conditional
ADE	adessive	APPR	apprehensive	CONJ	conjunction
ADJ	adjective	ART	article	COP	copula

DAT	dative	IND	indicative	PRO	pronoun
DEF	definite	INDF	indefinite	PROG	progressive
DEM	demonstrative	INE	inessive	PROH	prohibitive
DEMARC	demarcative	INF	infinitive	PROX	proximal
DET	determiner	INS	instrumental	PRS	present
DIM	diminutive	INT	intensifier	PST	past
DIR	direct	INTER	interessive	PTCP	participle
DIST	distal	INTR	intransitive	PTV	partitive
DS	different subject	INV	inverse	PURP	purposive
DU	dual	IPFV	imperfective	Q	question
DUR	durative	ITER	iterative	REC	recent
ELA	elative	JUS	jussive	RECP	reciprocal
EMP	emphatic	LOC	locative	REFL	reflexive
ERG	ergative	LOG	logophoric	REL	relative
ESS	essive	M	masculine	SBJ	subject
EXCL	exclusive	MID	middle	SBJV	subjunctive
EXIST	existential	N	neuter	SEQ	sequential
EXPER	experiential	NEG	negative	SG	singular
F	feminine	NFUT	nonfuture	SS	same subject
FOC	focus	NHON	nonhonorific	STAT	stative
FP	final particle	NN	non-neuter	SUBORD	subordinator
FUT	future	NOM	nominative	SUPL	superlative
GEN	genitive	NPST	nonpast	TERM	terminative
GER	gerund	NVIS	invisible	TOP	topical
HAB	habitual	OBJ	object	TR	transitive
HBL	humble	OBL	oblique	TRANSL	translative
HON	honorific	OBV	obviative	V	verb
HSY	hearsay	PASS	passive	VEG	vegetable
ILL	illative	PER	perlative	VEN	venitive
IMP	imperative	PFV	perfective	V _{INTR}	intransitive verb
IMPRS	impersonal	PL	plural	VIT	vital
INAL	inalienable	POSS	possessive	VOC	vocative
INAN	inanimate	POST	postposition	V _{TR}	transitive verb
INCHO	inchoative	POT	potential		
INCL	inclusive	PREP	prepositional		

Language codes

The language codes are from ISO 639-3 (SIL 2015).

AAR	Afar	ACV	Achumawi
ABK	Abkhaz	ADI	Adi
ABQ	Abaza	ADS	Adamorobe Sign Language
ACF	Saint Lucian Creole French	ADX	Amdo
ACU	Achuar-Shiwiar	ADY	Adyghe

AEB	Tunisian Arabic	ARK	Arikapú
AED	Argentine Sign Language	ARL	Arabela
AEN	Armenian Sign Language	ARN	Mapudungun
AER	Eastern Arrernte	ARP	Arapaho
AFN	Defaka	ARQ	Algerian Arabic
AFR	Afrikaans	ARR	Karo
AGR	Aguaruna	ARW	Arawak
AGU	Awakateko	ARY	Moroccan Arabic
AHK	Akha	ARZ	Egyptian Arabic
AIN	Ainu	ASB	Assiniboine
AKA	Akan	ASE	American Sign Language
AKE	Akawaio	ASF	Australian Sign Language
AKK	Akkadian	ASM	Assamese
AKZ	Alabama	ASP	Algerian Sign Language
ALC	Qawasqar	ASQ	Austrian Sign Language
ALE	Aleut	ASV	Asoa
ALN	Geg Albanian	ATB	Atsi
ALQ	Algonquin	ATX	Arutani
ALR	Alutor	AUC	Waorani
ALS	Tosk Albanian	AUE	ⱭKx'au 'ein
ALT	Altai	AUJ	Awjilah
ALY	Alyawarr	AVA	Avar
AMC	Amahuaca	AVE	Avestan
AME	Yanesha'	AWE	Awetí
AMH	Amharic	AYL	Libyan Arabic
AMI	Amis	AYO	Ayoreo
AMR	Amarakaeri	AYR	Aymara
AMU	Amuzgo	AZB	Azerbaijani
AMW	Western Aramaic	BAH	Bahamas Creole English
AMX	Anmatyerre	BAK	Bashkir
ANG	Old English	BAL	Balochi
ANI	Andi	BAM	Bambara
ANO	Andoque	BAN	Balinese
AOC	Pemon	BBL	Batsbi
AOI	Anindilyakwa	BCA	Bai
APC	North Levantine Arabic	BCI	Baoulé
APD	Sudanese Arabic	BCR	Babine
APJ	Jicarilla	BDQ	Bahnar
APM	Mescalero-Chiricahua	BEL	Belarusian
APU	Apurinã	BEN	Bengali
APW	Western Apache	BFI	British Sign Language
ARA	Arabic	BFT	Balti
ARB	Classical Arabic	BIN	Edo
ARC	Aramaic	BIS	Bislama
ARE	Western Arrernte	BLA	Blackfoot
ARI	Arikara	BLB	Bilua

BLC	Bella Coola	CEG	Chamacoco
BLK	Pa'ò	CES	Czech
BLT	Tai Dam	CHA	Chamorro
BOA	Bora	CHD	Highland Oaxaca Chontal
BOD	Central Tibetan	CHE	Chechen
BOG	Bamako Sign Language	CHF	Chontal
BOR	Borôro	CHJ	Ojtlán
BQY	Bengkala Sign Language	CHK	Trukese
BRE	Breton	CHL	Cahuilla
BRH	Brahui	CHM	Mari
BRX	Bodo	CHN	Chinook Jargon
BSK	Burushaski	CHO	Choctaw
BUA	Buriat	CHP	Chipewyan
BUG	Buginese	CHQ	Quiotepec
BUL	Bulgarian	CHR	Cherokee
BVL	Bolivian Sign Language	CHU	Old Church Slavonic
BWR	Bura	CHV	Chuvash
BZD	Bribri	CHY	Cheyenne
BZH	Buang	CIC	Chickasaw
BZJ	Belize Kriol	CIW	Chippewa
BZS	Brazilian Sign Language	CJA	Cham
CAB	Garifuna	CJO	Ashéninka Pajonal
CAC	Chuj	CKT	Chukchi
CAD	Caddo	CKU	Koasati
CAG	Nivaclé	CLC	Chilcotin
CAK	Kaqchikel	CLO	Lowland Oaxaca Chontal
CAO	Chácobo	CMI	Southern Emberá
CAP	Chipaya	CMN	Mandarin Chinese
CAQ	Car	CNI	Asháninka
CAR	Carib	CNL	Lalana-Tepinapa
CAS	Tsimané	COC	Cocopa
CAT	Catalan	COD	Cocama-Cocamilla
CAV	Cavineña	COJ	Cochimí
CAX	Chiquitano	COL	Columbian
CAY	Cayuga	COM	Comanche
CBA	Chibchan	CON	Cofán
CBI	Chachi	COO	Comox
CBK	Chavacano	COP	Coptic
CBN	Nyahkur	COR	Cornish
CBR	Cashibo-Cacataibo	CPA	Palantla
CBS	Kashinawa	CPC	Apurucayali
CBT	Chayahuita	CPI	Chinese Pidgin English
CBU	Candoshi-Shapra	CQD	Chuanqiandian
CDO	Eastern M'ń	CRB	Island Carib
CDS	Chadian Sign Language	CRD	Coeur d'Alene
CEB	Cebuano	CRE	Cree

CRG	Michif	ECS	Ecuadorian Sign Language
CRH	Crimean Tatar	EFE	Efe
CRK	Plains Cree	EFI	Efik
CRN	Cora	EGY	Egyptian
CRO	Crow	ELL	Greek
CRQ	Iyo'wujwa Chorote	EMK	Maninka
CRS	Seselwa	EMP	Northern Emberá
CRX	Carrier	EMS	Pacific Gulf Yupik
CSD	Chiangmai Sign Language	ENG	English
CSE	Czech Sign Language	ENH	Enets
CSF	Cuban Sign Language	ENL	Enlhet
CSG	Chilean Sign Language	ENM	Middle English
CSL	Chinese Sign Language	ENO	Enggano
CSN	Colombian Sign Language	ENQ	Enga
CSQ	Croatian Sign Language	EPO	Esperanto
CSR	Costa Rican Sign Language	ESE	Ese Ejja
CTO	Emberá-Catío	ESI	Northern Alaskan Iñupiaq
CTU	Chol	ESK	Seward Peninsula Iñupiaq
CUB	Cubeo	ESN	Salvadoran Sign Language
CUC	Usila	ESS	Central Siberian Yupik
CUK	San Blas Kuna	EST	Estonian
CUL	Kulina	ESU	Central Alaskan Yup'ik
CUX	Cuicatec	EUS	Basque
CYA	Nopala	EVE	Even
CYM	Welsh	EVN	Evenki
CZN	Zenzontepec	EWE	Ewe
CZO	Central M'ñ	FAO	Faroese
DAK	Dakota	FAS	Persian
DAL	Dahalo	FCS	Quebec Sign Language
DAN	Danish	FIA	Nobiin
DAR	Dargwa	FIJ	Fijian
DBL	Dyirbal	FIN	Finnish
DDO	Tsez	FLA	Kalispel-Salish
DEU	German	FNG	Fanagalo
DGR	Dogrib	FON	Fon
DIH	Kumiai	FRA	French
DIN	Dinka	FRO	Old French
DIS	Dimasa	FRY	Western Frisian
DIV	Dhivehi	FSE	Finnish Sign Language
DJK	Aukan	FSL	French Sign Language
DNW	Western Dani	FUC	Pulaar
DSE	Sign Language of the Netherlands	FUF	Pular
DTA	Dagur	FUL	Fula
DTS	Dogon	FUN	Iatê
DUJ	Dhuwal	FUR	Friulian
DZO	Dzongkha	FVR	Fur

GAB	Gabri	HAF	Haiphong Sign Language
GAG	Gagauz	HAK	Hakka
GAN	Gàn	HAT	Haitian Kreyòl
GAZ	West Central Oromo	HAU	Hausa
GCF	Guadeloupean Kreyòl	HAW	Hawaiian
GCR	Guianese Creole French	HAY	Haya
GDQ	Mehri	HBB	Huba
GEZ	Ge'ez	HBS	Serbo-Croatian
GIA	Kija	HCH	Huichol
GIL	Gilbertese	HDN	Haida
GIQ	Gelao	HDS	Honduran Sign Language
GIT	Gitxsan	HDY	Hadiyya
GLA	Scottish Gaelic	HEB	Hebrew
GLD	Nanai	HGM	Hai om
GLE	Irish	HID	Hidatsa
GLV	Manx	HIG	Kamwe
GMY	Mycenaean Greek	HIL	Hiligaynon
GNI	Gooniyandi	HIN	Hindi
GON	Gondi	HIT	Hittite
GOT	Gothic	HIX	Hixkaryána
GRC	Ancient Greek	HKS	Hong Kong Sign Language
GRN	Guaraní	HLU	Luwian
GRT	Garo	HMF	Hmong Don
GSE	Ghanaian Sign Language	HMN	Hmong
GSG	German Sign Language	HMO	Hiri Motu
GSM	Guatemalan Sign Language	HMS	Qiandong Miao
GSO	Gbaya	HNJ	Hmong Njua
GSS	Greek Sign Language	HOC	Ho
GTA	Guató	HOP	Hopi
GUB	Guajará	HOS	Ho Chi Minh City Sign Language
GUC	Wayuu	HPS	Hawaiian Sign Language
GUG	Paraguayan Guaraní	HRM	A-Hmao
GUH	Guahibo	HRV	Croatian
GUI	Eastern Bolivian Guaraní	HSH	Hungarian Sign Language
GUJ	Gujarati	HSL	Hausa Sign Language
GUK	Gumuz	HSN	Xiāng
GUL	Gullah	HTO	Minica Huitoto
GUN	Mbyá Guaraní	HTS	Hadza
GUP	Gunwinggu	HUB	Huambisa
GUU	Yanomamō	HUC	†Hua
GVC	Guanano	HUN	Hungarian
GWI	Gwich'in	HUP	Hupa
GYM	Ngäbere	HUR	Halkomelem
GYN	Guyanese Creole English	HUS	Huastec
GYR	Guarayu	HUU	Murui Huitoto
HAB	Hanoi Sign Language	HUV	Huave

HWC	Hawaii Creole	KAL	Kalaallisut
HYE	Armenian	KAN	Kannada
IBO	Igbo	KAQ	Capanahua
ICR	Islander Creole English	KAS	Kashmiri
IDB	Indo-Portuguese	KAT	Georgian
IGN	Ignaciano	KAU	Kanuri
III	Yi	KAY	Kamayurá
IJC	Izon	KAZ	Kazakh
IKE	Inuktitut	KBD	Kabardian
IKT	Inuinnaqtun	KBH	Camsá
ILO	Ilocano	KBL	Kanembu
INA	Interlingua	KBY	Manga Kanuri
IND	Bahasa Indonesia	KCA	Khanty
INH	Ingush	KCG	Katab
INL	Indonesian Sign Language	KCR	Katla
INS	Indo-Pakistani Sign Language	KDT	Kuy
IQU	Iquito	KEE	Keres
IRN	Irántxe	KEK	Q'eqchi'
ISE	Italian Sign Language	KET	Ket
ISG	Irish Sign Language	KFA	Kodava
ISL	Icelandic	KFB	Kolami
ITA	Italian	KFE	Kota
ITL	Itelmen	KFQ	Korku
ITO	Itonama	KGG	Kusunda
IUM	Iu Mien	KGK	Kaiwá
IXL	Ixil	KHA	Khasi
JAC	Jakalteko	KHG	Khams
JAM	Jamaican Creole	KHM	Khmer
JAV	Javanese	KIB	Koalib
JBE	Judeo-Berber	KIC	Kickapoo
JBO	Lojban	KIK	Gikuyu
JBT	Jabutí	KIN	Kinyarwanda
JCS	Jamaican Country Sign Language	KIO	Kiowa
JEB	Jebero	KIR	Kyrgyz
JIC	Tol	KJB	Q'anjob'al
JIV	Shuar	KJD	Southern Kiwai
JLS	Jamaican Sign Language	KJG	Khmu
JOS	Jordanian Sign Language	KJJ	Xinalug
JPN	Japanese	KJP	Eastern Pwo
JQR	Jaqaru	KJU	Kashaya
JSL	Japanese Sign Language	KKY	Guugu Yimidhirr
JUN	Juang	KMC	Dong
JUR	Jurúna	KNC	Central Kanuri
KAA	Karakalpak	KNJ	Akateko
KAB	Kabyle	KNM	Kanamari
KAC	Jingpho	KNT	Panoan Katukína

KNW	Kung-Ekoka	LIL	Lillooet
KOG	Kogi	LIR	Liberian Pidgin English
KOM	Komi	LIS	Lisu
KON	Kongo	LIT	Lithuanian
KOR	Korean	LKT	Lakota
KPC	Curripaco	LLD	Ladin
KPJ	Karajá	LLS	Lithuanian Sign Language
KPY	Koryak	LOU	Louisiana Creole
KRC	Karachai-Balkar	LOY	Loke
KRI	Krio	LSG	Lyons Sign Language
KRL	Karelian	LSI	Lashi
KRU	Kurukh	LSP	Panamanian Sign Language
KSB	Shambala	LTC	Middle Chinese
KSD	Tolai	LUG	Ganda
KSW	S'gaw	LUI	Luiseño
KTN	Karitiána	LUO	Dholuo
KTU	Kituba	LUS	Mizo
KTX	Kaxararí	LUT	Lushootseed
KTZ	Ju'hoan	LZZ	Laz
KUM	Kumyk	MAA	San Jerónimo Tecóatl
KUN	Kunama	MAD	Madurese
KUR	Kurdish	MAF	Mafa
KUT	Kutenai	MAH	Marshallese
KVK	Korean Sign Language	MAJ	Jalapa de Díaz
KWI	Awa-Cuaiquer	MAL	Malayalam
KWK	Kwakiutl	MAM	Mam
KXO	Kanoé	MAR	Marathi
KXU	Kui	MAS	Maasai
KYH	Karuk	MAU	Huautla
KYU	Red Karen	MAV	Sateré-Mawé
KYZ	Kayabí	MAZ	Mazahua
LAC	Lacandón	MBC	Macushi
LAD	Judeo-Spanish	MBL	Maxakalí
LAO	Lao	MCB	Machiguenga
LAQ	Qabiao	MCD	Sharanahua
LAT	Latin	MCF	Matsés
LAV	Latvian	MCH	Maquiritari
LBE	Lak	MCN	Masa
LBJ	Ladakhi	MDE	Maba
LBZ	Lardil	MDF	Moksha
LDN	Láadan	MDJ	Mangbetu
LEP	Lepcha	MDL	Maltese Sign Language
LES	Lese	MEN	Mende
LEZ	Lezgi	MEU	Motu
LHU	Lahu	MEZ	Menominee
LIC	Hlai	MFE	Morisyen

MFS	Mexican Sign Language	MSD	Yucatec Maya Sign Language
MFY	Mayo	MSE	Musey
MGD	Moru	MSR	Mongolian Sign Language
MHC	Mocho	MTO	North Highland Mixe
MHD	Mbugu	MTS	Yora
MHI	Ma'di	MUB	Mubi
MHJ	Mogholi	MUD	Mednyj Aleut
MHQ	Mandan	MUE	Media Lengua
MHX	Maru	MUS	Muskogee
MIC	Micmac	MVI	Miyako
MID	Mandaic	MWF	Murrinh-Patha
MIK	Mikasuki	MWP	Kala Lagaw Ya
MIQ	Mískito	MWT	Moken
MIR	Lowland Mixe	MWW	Hmong Daw
MJC	San Juan Colorado	MXP	Zempoaltepetl
MJG	Tu	MXQ	South Midland Mixe
MJT	Malto	MXT	Jamiltepec
MKD	Macedonian	MXV	Metlatónoc
MKM	Moklen	MYA	Burmese
MKS	Silacayoapan	MYP	Pirahã
MKZ	Makasae	MYU	Mundurukú
MLG	Malagasy	MYV	Erzya
MLS	Masalit	MZP	Movima
MLT	Maltese	MZR	Marúbo
MMR	Xiangxi Miao	MZS	Macanese
MNC	Manchu	NAB	Southern Nambikuára
MND	Mondé	NAG	Naga Pidgin
MNP	Northern M'ñ	NAN	Southern M'ñ
MNR	Mono	NAQ	Nama
MNS	Mansi	NAV	Navajo
MNW	Mon	NBE	Konyak
MOC	Mocoví	NGC	Nisga'a
MOD	Mobilian Jargon	NCI	Classical Nahuatl
MOH	Mohawk	NCL	Michoacán Nahuatl
MON	Mongolian	NCS	Nicaraguan Sign Language
MOS	Mooré	NEP	Nepali
MOV	Mojave	NEW	Newari
MPG	Marba	NEZ	Nez Perce
MPM	Yosondúa	NGA	Ngbaka
MRC	Maricopa	NHE	Eastern Huasteca Nahuatl
MRE	Martha's Vineyard Sign Language	NIO	Nganasan
MRG	Miri	NIT	Naiki
MRI	Maori	NIV	Nivkh
MRT	Margi Central	NJM	Angami
MRW	Maranao	NJO	Ao
MSA	Malay	NJZ	Nyishi

NLD	Dutch	PAV	Wari'
NMF	Tangkhul	PAW	Pawnee
NMU	Maidu	PBB	Páez
NNC	Nancere	PBC	Patamona
NON	Old Norse	PBH	E'ñapa Woromaipu
NOR	Norwegian	PBS	Central Pame
NOT	Nomatsiguenga	PCA	Western Popoloca
NSI	Nigerian Sign Language	PCI	Duruwa
NSM	Sema	PCM	Nigerian Pidgin English
NSO	Northern Sotho	PEI	Chichimeca-Jonaz
NSP	Nepalese Sign Language	PEO	Old Persian
NSR	Maritime Sign Language	PGA	Juba Arabic
NTP	Northern Tepehuán	PHA	Pa-Hng
NUK	Nootka	PHN	Phoenician
NUP	Nupe	PIA	Pima Bajo
NUS	Nuer	PIB	Yine
NXQ	Naxi	PID	Piaroa
NYH	Nyigina	PIO	Piapoco
NZS	New Zealand Sign Language	PIS	Pijin
OCH	Old Chinese	PJT	Pitjantjatjara
OCI	Occitan	PLG	Pilagá
OJI	Ojibwa	PLN	Palenquero
OJP	Old Japanese	PLO	Oluta Popoluca
OKA	Okanagan	PLU	Palikúr
OMA	Omaha-Ponca	PML	Lingua Franca
ONB	Ongbe	PMQ	Northern Pame
ONE	Oneida	POH	Poqomchi'
ONO	Onondaga	POI	Sierra Popoluca
ONW	Old Nubian	POL	Polish
OOD	O'odham	POR	Portuguese
OPY	Ofayé	POS	Sayula Popoluca
ORM	Oromo	POT	Potawatomi
OSC	Oscan	POV	Crioulo
OSS	Ossetic	PPI	Paipai
OTB	Old Tibetan	PPL	Pipil
OTE	Mezquital	PQM	Malecite-Passamaquoddy
OTM	Eastern Highland	PRG	Old Prussian
OTN	Tenango	PRK	Parauk
OTQ	Querétaro	PRL	Peruvian Sign Language
OTT	Temoaya	PRS	Dari
OTW	Ottawa	PRX	Purik
OYM	Wayampi	PRZ	Providencia Sign Language
PAB	Parecís	PSC	Persian Sign Language
PAH	Tenharim	PSD	Plains Indian Sign Language
PAN	Punjabi	PSG	Penang Sign Language
PAP	Papiamentu	PSL	Puerto Rican Sign Language
PAU	Palau	PSO	Polish Sign Language

PSP	Philippine Sign Language	SGW	Central West Gurage
PSR	Portuguese Sign Language	SHH	Shoshone
PUA	Western Highland Purepecha	SHI	Tachelhit
PUI	Puinave	SHN	Shan
PUS	Pashto	SHP	Shipibo-Conibo
PWI	Patwin	SHS	Shuswap
PWN	Paiwan	SID	Sidamo
PYE	Pye Krumen	SIN	Sinhala
PYS	Paraguayan Sign Language	SJO	Xibe
QUC	K'iche'	SJW	Shawnee
QUE	Quechua	SKD	Miwok
QUG	Chimborazo Highland Quichua	SLC	Sáliba
QUH	South Bolivian Quechua	SLK	Slovak
QUY	Ayacucho Quechua	SLR	Salar
QUZ	Cusco Quechua	SLS	Singapore Sign Language
QXS	Southern Qiang	SLV	Slovenian
RAP	Rapa Nui	SMA	Southern Sami
RAS	Tegali	SME	Northern Sami
RCF	Réunion Creole French	SMO	Samoan
REJ	Rejang	SNA	Shona
RIF	Tarifit	SOM	Somali
RKB	Rikbaktsa	SON	Songhai
ROH	Romansch	SOT	Sotho
ROM	Romani	SPA	Spanish
RON	Romanian	SPC	Sapé
ROO	Rotokas	SPO	Spokane
ROP	Kriol	SPT	Spiti Bhoti
RSL	Russian Sign Language	SQI	Albanian
RTM	Rotuman	SQK	Albanian Sign Language
RUN	Rundi	SQS	Sri Lankan Sign Language
RUS	Russian	SQU	Squamish
RYU	Okinawan	SRB	Sora
SAC	Fox	SRD	Sardinian
SAD	Sandawe	SRM	Saramaccan
SAE	Sabanè	SRN	Sranan
SAG	Sango	SRP	Serbian
SAH	Yakut	SSP	Spanish Sign Language
SAN	Sanskrit	SSW	Swazi
SAT	Santali	STO	Stoney
SCI	Sri Lankan Creole Malay	STP	Southern Tepehuán
SCO	Scots	STR	Straits Salish
SDL	Saudi Arabian Sign Language	STV	East Gurage
SEC	Sechelt	SUN	Sundanese
SEE	Seneca	SVA	Svan
SEI	Seri	SVK	Slovakian Sign Language
SEL	Selkup	SWA	Swahili
SFS	South African Sign Language	SWE	Swedish

SWI	Sui	TOJ	Tojolabal
SWL	Swedish Sign Language	TOL	Tolowa
SWN	Sawknah	TON	Tongan
SYC	Syriac	TOO	Northern Totonac
SYW	Kagate	TOP	Papantla
SYZ	Al-Sayyid Bedouin Sign Language	TOS	Sierra Totonac
TAB	Tabassaran	TOW	Jemez
TAE	Tariana	TPI	Tok Pisin
TAH	Tahitian	TPJ	Ñandeva
TAM	Tamil	TPP	Pisaflores
TAQ	Timbuktu Tamasheq	TPR	Tuparí
TAR	Tarahumara	TPT	Tlachichilco
TAT	Volga Tatar	TPW	Tupí
TAY	Atayal	TPX	Tlapanec
TBA	Aikanā	TPY	Trumai
TCA	Ticuna	TQN	Tenino
TCS	Torres Strait Creole	TRF	Trinidadian Creole English
TCX	Toda	TRN	Trinitario
TCY	Tulu	TRU	Turoyo
TCZ	Thado	TRV	Seediq
TEA	Temiar	TSE	Tunisian Sign Language
TED	Tepo Krumen	TSI	Tsimshian
TEE	Huehuetla	TSM	Turkish Sign Language
TEL	Telugu	TSN	Tswana
TER	Terêna	TSQ	Thai Sign Language
TET	Tetum	TSS	Taiwanese Sign Language
TEW	Tewa	TSU	Tsou
TGK	Tajik	TSY	Tebul Sign Language
TGL	Tagalog	TTQ	Tawallammat Tamajaq
THA	Thai	TUB	Tübatulabal
THP	Thompson	TUK	Turkmen
THV	Tahaggart Tamahaq	TUO	Tucano
THZ	Tayart Tamajeq	TUR	Turkish
TIG	Tigré	TUS	Tuscarora
TIR	Tigrinya	TWF	Northern Tiwa
TIV	Tiv	TXB	Tocharian B
TIW	Tiwi	TXG	Tangut
TIX	Southern Tiwa	TYV	Tuvan
TLH	Klingon	TZH	Tzeltal
TLI	Tlingit	TZJ	Tz'utujil
TLP	Filomeno Mata-Coahuilán	TZM	Tamazight
TMF	Toba-Mascoy	TZO	Tzotzil
TMH	Tuareg	UBY	Ubykh
TNA	Tacana	UDM	Udmurt
TNQ	Taino	UGY	Uruguayan Sign Language
TOB	Toba	UIG	Uyghur
TOC	Coyutla Totonac	UKR	Ukrainian

UKS	Kaapor Sign Language	XTG	Gaulish
ULK	Meriam	XTN	Northern Tlaxiaco
UMA	Umatilla	XTO	Tocharian A
UNR	Mundari	XUM	Umbrian
URA	Urarina	XUU	Khwe
URB	Kaapor	YAA	Yaminahua
URD	Urdu	YAD	Yagua
URE	Uru	YAE	Pumé
UTE	Ute-Chemehuevi	YAK	Yakima
UZN	Uzbek	YAN	Mayangna
VAR	Guarijío	YAP	Yapese
VIE	Vietnamese	YAQ	Yaqui
VMC	Juxtlahuaca	YAU	Yuwana
VSL	Venezuelan Sign Language	YDD	Yiddish
WAA	Walla Walla	YKG	Northern Yukaghir
WAC	Kiksht	YNK	Naukan Yupik
WAL	Wolaytta	YOK	Yokuts
WAP	Wapishana	YOR	Yoruba
WAR	Waray-Waray	YRK	Nenets
WAS	Washo	YRL	Nheengatú
WAW	Waiwai	YUA	Yucatec
WAY	Wayana	YUC	Yuchi
WBA	Warao	YUE	Yuè
WBP	Warlpiri	YUF	Upland Yuman
WEN	Sorbian	YUM	Quechan
WES	Cameroonian Pidgin English	YUP	Yukpa
WIC	Wichita	YUR	Yurok
WIN	Winnebago	YUX	Southern Yukaghir
WIY	Wiyot	YUZ	Yuracare
WLV	Wichí Lhamtés Vejoz	ZAC	Ocotlán
WOL	Wolof	ZAI	Isthmus
WTI	Berta	ZAM	Miahuatlán
WUU	Wú	ZAP	Zapotec
XAL	Oirat	ZAU	Zangskari
XAV	Xavánte	ZEN	Zenaga
XAW	Kawaiisu	ZHO	Chinese
XCT	Classical Tibetan	ZOC	Central Zoque
XER	Xerénte	ZOH	Santa María Chimalapa Zoque
XFA	Faliscan	ZOR	Northeast Zoque
XHO	Xhosa	ZOS	North Zoque
XLC	Lycian	ZPC	Choapan
XML	Malaysian Sign Language	ZPK	Tlacolulita
XNZ	Kenzi	ZRO	Záparo
XOK	Xokleng	ZUL	Zulu
XPE	Kpelle	ZUN	Zuni
XSL	Slavey	ZYB	Zhuang
XSR	Sherpa		

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2005)

CONSONANTS (PULMONIC)

© 2005 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

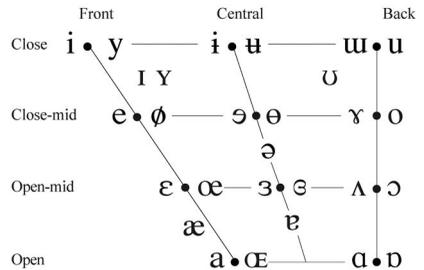
CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
ɸ Bilabial	ɓ Bilabial	ʼ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ɟ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɡ Velar	kʼ Velar
ǁ Alveolar lateral	ɠ Uvular	sʼ Alveolar fricative

OTHER SYMBOLS

- ɱ** Voiceless labial-velar fricative
 - ɰ** Voiced labial-velar approximant
 - ɸ** Voiced labial-palatal approximant
 - ħ** Voiceless epiglottal fricative
 - ʕ** Voiced epiglottal fricative
 - ʡ** Epiglottal plosive
 - ɕ ʑ** Alveolo-palatal fricatives
 - ɻ** Voiced alveolar lateral flap
 - ɥ** Simultaneous **ɥ** and **X**
- Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

- ˈ** Primary stress
 - ˌ** Secondary stress
 - ː** Long
 - ˑ** Half-long
 - ˑ̇** Extra-short
 - ̤** Minor (foot) group
 - ̥** Major (intonation) group
 - Syllable break
 - ˘** Linking (absence of a break)
- Example: **ˈkʰɒnəˈtʃən**

DIACRITICS Diacritics may be placed above a symbol with a descender, e.g. **ɨ̥**

◌ [◌] Voiceless	ɲ̥ ɖ̥	◌ ^{◌̤} Breathy voiced	ɲ̤ ɖ̤	◌ ^{◌̥} Dental	t̥ d̥
◌ ^{◌̤} Voiced	ɲ̤ ɖ̤	◌ ^{◌̥} Creaky voiced	ɲ̥ ɖ̥	◌ ^{◌̥} Apical	t̥̚ d̥̚
◌ ^{◌̥} Aspirated	t̥ʰ d̥ʰ	◌ ^{◌̥} Linguolabial	t̥̚ ɖ̥̚	◌ ^{◌̥} Laminar	t̥̚̚ d̥̚̚
◌ ^{◌̥} More rounded	ɔ̙	◌ ^{◌̥} Labialized	t̥ʷ ɖ̥ʷ	◌ ^{◌̥} Nasalized	ẽ
◌ ^{◌̥} Less rounded	ɔ̜	◌ ^{◌̥} Palatalized	t̥ʲ ɖ̥ʲ	◌ ^{◌̥} Nasal release	d̥ⁿ
◌ ^{◌̥} Advanced	ɥ̟	◌ ^{◌̥} Velarized	t̥˞ ɖ̥˞	◌ ^{◌̥} Lateral release	d̥^l
◌ ^{◌̥} Retracted	ɥ̠	◌ ^{◌̥} Pharyngealized	t̥ˤ ɖ̥ˤ	◌ ^{◌̥} No audible release	d̥^ɹ
◌ ^{◌̥} Centralized	ẽ	◌ ^{◌̥} Velarized or pharyngealized	ɟ̠		
◌ ^{◌̥} Mid-centralized	ẽ̞	◌ ^{◌̥} Raised	e̞ (ɹ̞ = voiced alveolar fricative)		
◌ ^{◌̥} Syllabic	ɲ̩	◌ ^{◌̥} Lowered	e̝ (β̝ = voiced bilabial approximant)		
◌ ^{◌̥} Non-syllabic	ɲ̩̥	◌ ^{◌̥} Advanced Tongue Root	e̟		
◌ ^{◌̥} Rhoticity	ɻ̥ ɑ̞̥	◌ ^{◌̥} Retracted Tongue Root	e̠		

- TONES AND WORD ACCENTS
- | LEVEL | CONTOUR |
|----------------------------------|------------------------------|
| ẽ̥ or ˥ Extra high | ẽ̥ or ˥ Rising |
| é High | ẽ̂ Falling |
| ē Mid | ẽ̃ High rising |
| è Low | ẽ̄ Low rising |
| ẽ̇ Extra low | ẽ̅ Rising-falling |
| ˩ Downstep | ↗ Global rise |
| ˨˨˩ Upstep | ↘ Global fall |

FIGURE 0.1 IPA chart, <http://www.internationalphoneticassociation.org/content/ipa-chart>, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.

Phonetic symbols

The symbols in phonetic and phonemic transcriptions are from the alphabet of the International Phonetic Association (IPA), as explained in Figure 0.1.

In the table labeled “CONSONANTS (PULMONIC)”, the column headers indicate place of articulation, and the row labels indicate manner of articulation. In the table labeled “VOWELS”, the column headers indicate how far the dorsum of the tongue is moved toward the front or back of the mouth and the row labels indicate how close the dorsum is to the roof of the mouth.

An Introduction to the Languages of the World

Classification of Languages

Why bother classifying languages?

Comparing and contrasting the thousands of languages in the world would be overwhelming without an efficient framework for grouping them in a reasonably small number of categories. But what are the appropriate categories? Sorting language names in alphabetical order is useful for some purposes, such as quickly looking up a language in a book index, but no one would see any point to devoting a year to comparing all languages whose name begins with the letter ⟨G⟩. In contrast, classifying languages by more essential properties can efficiently reveal their most important similarities and differences. This in turn yields information about the nature of language itself – the primary goal of linguistics.

This chapter investigates the most common ways to classify languages. Each is based on a different set of important language properties. A **GENETIC** classification approaches languages historically, grouping them together if they share a common ancestral language. Classifying languages by historical origin takes a step toward reconstructing what the ancestral languages were like and is useful for exploring how languages change over time. A **TYPOLICAL** approach groups languages according to similarities in linguistic structure. Current typologies look at aspects of syntax, morphology, or phonology.

1.1 Genetic classification

Genetic classification organizes languages by their parentage: which language did a language evolve from? It comprises an analysis of anagenesis and cladogenesis.

1.1.1 ANAGENESIS

It has always been obvious that certain sets of languages are modified forms of languages that are well known from the historical record. For example, English as spoken today is clearly a modified form of the language Chaucer spoke around 1380,

which is a modified form of the language Cædmon spoke around 680. Indeed, we now know that INNOVATION is universal. All languages have evolved from earlier forms of the same language as innovations have arisen, spread to other speakers of the same language, and persisted as new generations of children learned the somewhat changed language. Accordingly, the first principle of genetic classification is to group DESCENDANT languages, such as today's English, with their ANCESTOR languages, such as Chaucer's or Cædmon's English.

Forms of a language that evolved from each other in a straight, unbranching line of descent, or ANAGENESIS, are referred to by the same name, although they may be qualified by terms such as *Old*, *Middle*, and *Modern*. Thus the language of Cædmon, Old English (ANG), is grouped with the language of Chaucer, Middle English (ENM), and both are grouped with our Modern English (ENG). (The three-letter codes in small caps are unique standard identifiers for the language; see the introduction to Chapter 3 and the list of language codes in the Symbols and Abbreviations section at the front of the book.) Other languages that have a sufficiently long documented history may be similarly qualified, such as Old Japanese (OJP), Middle Japanese, and Modern Japanese (JPN). The relationships between the various forms of a language can be diagrammed as in Figure 1.1. The labeled ovals refer to languages, and the lines refer to direction of descent. In linguistics, there is a tradition of having time flow from the top of the page downward, so the line from Old English to Middle English means that Old English is the PARENT of Middle English, or, equivalently, that Middle English is the DAUGHTER of Old English. We use the words *parent* and *daughter* only when talking about adjacent languages connected by one line. As in the previous paragraph, we use the words *ancestor* and *descendant* to refer to the same concepts transitively. Thus Modern English is a descendant of Middle English and of Old English; Old English is an ancestor of Middle English and Modern English.

Importantly, genetic classification does not group together all Old languages, nor all Middle languages, nor all Modern languages. That is why Figure 1.1 has no connections at all drawn between the English lineage on the left and the Japanese lineage on the right. Languages really do not have anything at all in common just because they are called *Old* or just because they were spoken centuries or

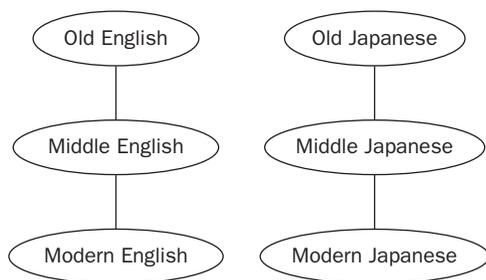


FIGURE 1.1 Anagenesis

millennia ago, other than some obvious vocabulary differences such as not having a word for ‘spaceship’ or ‘Facebook’. As far back as we can observe, languages have always worked in the same ways that they do today and were no better nor worse than modern languages by any objective measure. This idea is known as the UNIFORMITARIAN PRINCIPLE.

In contrast, languages in the same lineage, such as Old, Middle, and Modern English, tend to have many properties in common. Even though Old English may sound like a foreign language to speakers of Modern English, a very large number of words and much of the grammar are still recognizable; consider such words as *hand* ‘hand’, *scip* ‘ship’, and *cwen* ‘queen’. The familiarity of old vocabulary can be increased by an understanding of SOUND CHANGES that occur regularly throughout the vocabulary. For example, at the end of Middle English, the sound [u:] diphthongized, ending up as [aʊ] in present-day Modern English. Knowing that one rule, Middle English words like [hu:s], [nu:], and [u:t] become instantly recognizable as the precursors of *house*, *now*, and *out*, respectively. Sound changes do not help at all in connecting Old Japanese vocabulary with that of Old English.

When discussing relations between older and newer forms of a language, it is helpful to distinguish between innovations and RETENTIONS. An innovation is the product of a change; a retention is the ancestral state without any change. For example, some dialects in the north of England didn’t participate in the sound change of [u:] > [aʊ]. Other dialects did have some sort of change, but they differed in the exact output: [aʊ], [æʊ], [əʊ], and so forth. In discussing how words are pronounced in the different dialects, it may be convenient to refer to pronunciations such as [hu:s] as retentions, and pronunciations such as [haus], [hæʊs], and [həʊs] as innovations.

It is helpful to keep in mind that innovations occur rather frequently, yet not so frequently that grandparents cannot communicate with their grandchildren. You might imagine that every time an innovation spreads through a language, the language is different and so should be given a different name; but, in fact, hundreds or thousands of changes may take place during the documented lineage of a language, which would make it impractical to give a new label to the language every time it changes. In practice, the dividing lines between languages like Old versus Middle English are drawn quite arbitrarily. Sometimes they are drawn simply in order to make the time spans of different languages equivalent; the dates given for Middle English, 1100 to 1500, were chosen mostly so that Old and Middle English would both span 400 years. Sometimes the demarcations correspond to major cultural or historical events that are believed to have a large impact on the language. What is certainly not the case is that in the year 1100, Old English changed so much that it became a new language, Middle English. Figures such as 1.1 must not be read as implying clean breaks as one language changes into another, nor even as implying that English and Japanese had exactly three clearly demarcated stages.

1.1.1.2 CLADOGENESIS

Innovations sometimes spread through only part of a speech community. Barriers to the spread of innovation may be something as concrete as a river or a mountain range, but often they are intangible social or political factors. Wherever people feel weaker allegiance to those who originated an innovation than to those who have not adopted an innovation, that innovation may be stopped dead in its tracks. Conversely, innovations may spread through part of a speech community because people feel special allegiance to the group that originated the innovation.

The ultimate result of incomplete spread of innovations can be CLADOGENESIS: one language splits into descendant languages that are SISTERS. Scholars must have been aware of certain cases of cladogenesis for centuries. The long tradition of using classical Latin as the language of scholarship, religion, and law through much of Europe made people aware of the fact that their native languages were altered forms of Latin. The adoption of different sets of innovations in various parts of the Latin-speaking Roman Empire had led to much cladogenesis, with Latin developing into dozens of different languages, including Spanish, Portuguese, French, Italian, and Romanian: the Romance languages. By the late 18th century, linguists began to understand that cladogenesis must be a very common occurrence, and they hypothesized that similar languages arose through the splitting of a common ancestor. They were undeterred when there was no direct historical evidence of that ancestor language, because they knew that the vast majority of languages were never written down. They simply made up a name for the hypothesized ancestor, regularly beginning the name with *proto-*, a Greek root meaning ‘first’.

Cladogenesis can be represented by drawing TREES. A tree is a type of diagram much like that used for representing anagenesis (Figure 1.1), with the addition of one feature: a parent can have more than one child. Lines drawn from one parent language to two child languages mean that the parent split into two separate languages, at least one of which is characterized by innovations that did not spread throughout the whole parent language. Figure 1.2 is a tree that presents in simplified form one hypothesis of how the Germanic languages developed through cladogenesis. It says that there was once a language we call *Proto-Germanic*, which split up into a language we call *Proto-Northwest Germanic* and one we call *Proto-East Germanic*. *Proto-Northwest Germanic* in turn split into *Proto-West Germanic* and *Proto-North Germanic*, and so forth.

The term CLADE is used to refer to a language and all its descendants. For example, in Figure 1.2, *Proto-Northwest Germanic* and its children *Proto-West Germanic* and *Proto-North Germanic*, as well as all of their children down to their modern languages German, Yiddish, Dutch, and so forth, constitute a clade. When a clade descends from a language whose name starts with *proto-*, the clade can be conveniently named by dropping the prefix: the clade just described is called *Northwest Germanic*. The term BRANCH refers to a clade from the perspective of its parent; thus *Northwest Germanic* is a branch of the *Germanic* clade. In

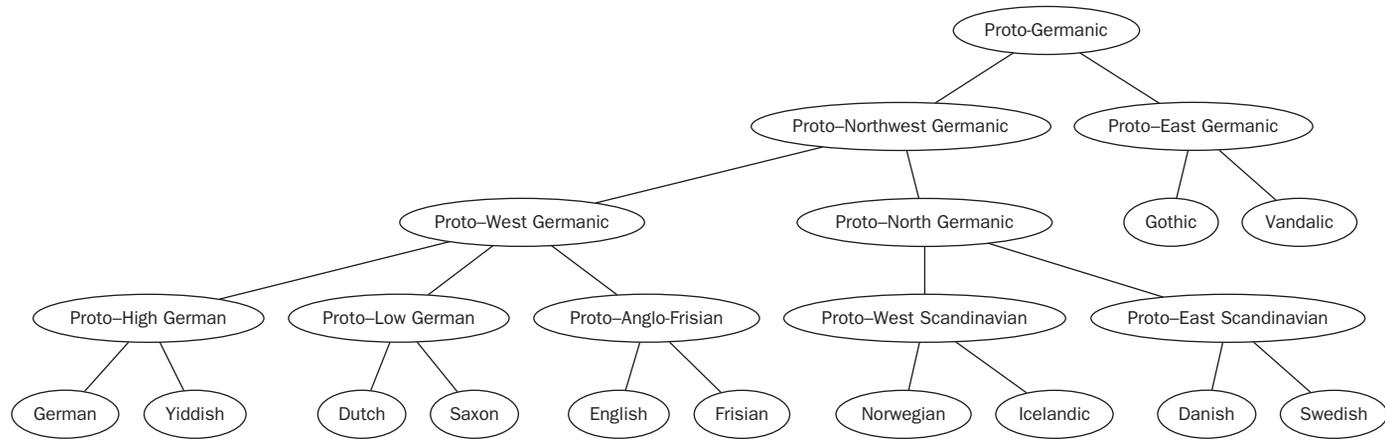


FIGURE 1.2 Cladogenesis

order to save space, Figure 1.2 does not show the parent of Proto-Germanic, but Germanic is a branch of Indo-European. The word *clade* is a bit more general; in particular, individual languages are clades, and so is the entire tree. Within a particular tree, a language is said to be more closely RELATED to a language with which it shares a nearer common ancestor than to one with which it shares a more remote ancestor. Thus English is more closely related to Frisian than it is to German, because Proto-Anglo-Frisian is a nearer ancestor of English than Proto-West Germanic is.

Cladistic analysis of languages is very useful because it maps out the patterns of SHARED INNOVATIONS. The splitting of any particular language is occasioned by innovations that are found in only one of the children. In the case of the split of Proto-Germanic, one of the innovations adopted by Proto-Northwest Germanic was a sound change that lowered *[e:] to *[a:]. (The asterisks indicate that the sound is RECONSTRUCTED, or inferred by linguistic reasoning, not directly attested in any text.) Thus Proto-Germanic *[le:tan] ‘to let’ became *[la:tan] in Proto-Northwest Germanic. All descendants of Proto-Northwest Germanic would be expected to share that innovation. Thus, Proto-North Germanic and Proto-West Germanic both retained the form *[la:tan], and they in turn passed the vowel *[a:] down to their children. Knowing that a language is in the Northwest Germanic clade immediately allows you to draw conclusions about the innovations it shares.

While a cladistic analysis helps us to understand the patterns of innovations, it also reveals the patterns of retentions. If Proto-Northwest Germanic is characterized by the lowering of *[e:] to *[a:], then we can conclude that Proto-East Germanic retained *[e:], and indeed the Proto-East Germanic word for ‘let’ was *[le:tan]. Further, absent innovation, we can conclude that all languages in a clade presumably retain the properties of their shared ancestor. The same conclusion can be drawn from anagenesis: later forms of a language tend to retain many properties of older forms of a language. Consequently, it is reasonable to infer that languages that are closely related to each other are more similar than languages that are more remotely related to each other, though especially high rates of innovation within a clade can bring about exceptions to this rule. Linguists overwhelmingly consider genetic classification, which combines analysis of cladogenesis and anagenesis, to be the single most useful way of grouping languages, largely because it is an objective way of grouping together languages that are most likely to share a wide range of properties. It is the main approach that we adopt in this book, although we often use geography as a way of organizing discussion of languages that have no known relation to each other.

1.1.3 PROBLEMS WITH GENETIC CLASSIFICATION

By this point you may have noticed that grouping languages in terms of their ancestors and descendants is very similar to how biologists group species in terms of their evolution. There is one major difference, however. In biology, there is ample

evidence that all species are related and that a cladistic analysis could, in principle, result in one huge tree that includes all species back to the very beginning of life on Earth. That is, virtually all biologists accept the theory of MONOGENESIS, that all species share a common ancestor species. In linguistics, it is far more difficult to choose between monogenesis and POLYGENESIS, the theory that languages may have developed from multiple unrelated ancestral languages. Therefore, even historical linguists who accept the monogenesis hypothesis agree that we should not simply assume languages to be related, but that we should prove it in all cases. Languages that can be proved to be related are said to be members of the same language FAMILY; languages that are not provably related are classified as members of different families. Put another way, the family is the maximal clade in linguistics. Some families, such as Indo-European, have hundreds of languages; some families, such as Basque, consist of a single language that cannot be proved to be related to any other language. It should be noted that this usage is rather different from how the word *family* is used in biology, where a family can be part of an order, class, phylum, and so forth, higher groupings for less-closely related species. In linguistics, there is nothing provably related above the family.

In our current state of knowledge, there are hundreds of language families, clades not proved to be related to each other. It is natural to feel a certain dissatisfaction with this number. Especially if one feels that language monogenesis is likely – that all languages descend ultimately from a single ancestor language – then being unable to prove the relatedness of hundreds of clades comes across as a serious failure. The problem, as linguists widely agree, is that languages innovate rapidly. Each change removes one bit of evidence that a given language is related to others. As changes accumulate over centuries and millennia, we are left with insufficient evidence to prove genetic relationships that may have existed at one time. The first languages must have appeared at least 50,000 years ago, yet 10,000 years is probably an upper limit to our ability to prove a genetic relationship, and the limit is even less in cases where the changes over time have been more drastic. Despite this, many scholars of linguistics and cladistics have strived to show that various families are related to each other. Occasionally they have used other terms, such as *phylum*, for speculative groupings larger than the family. Others have tried grouping families by their similarity to each other, on the theory that because related languages tend to be relatively similar to each other, the converse should hold: the more similar languages are, the more related they must be. Greenberg (1963; 1971; 1987) used similarity comparisons to group together massive numbers of language families. Bengtson & Ruhlen (1994) even claimed to have reconstructed several words of the original human language. It certainly would be interesting to bring all the world's languages into one family, and perhaps some of the proposals for larger families will one day be found to be on the right track. However, we are not persuaded by many of these ambitious proposals, and in this book we have preferred to err on the conservative side, using only proposals that are widely accepted by the experts in the language families in question.

At the other taxonomic extreme, it is common to use the term DIALECT to refer to a grouping smaller than a language. Basque, for example, is described as having five main dialects, which all differ among themselves. There is, unfortunately, no objective linguistic definition that lets us tell whether a clade comprises different languages or dialects of the same language. The most widespread suggestion is that only dialects are MUTUALLY INTELIGIBLE, that is, their speakers can understand each other. That standard is surprisingly difficult to put to practical use, however, because intelligibility is a continuum that depends on many factors. You should not be surprised or offended if we refer to two speech forms as different dialects when some other book calls them different languages, or vice versa; there simply is no way to distinguish those terms.

Another terminological issue with genetic classification is that it is very easy to confuse language genetics with human GENETICS. Most emphatically, linguists do not classify languages by analyzing the DNA of the people who speak them. It is true that people tend to marry people who speak the same language and raise their children to speak that same language; that leads to a correlation between human genetics and language genetics. That correlation may be interesting, but there can be no correlating between linguistics and biology unless linguists do their own job first and provide fully independent analyses of language evolution based on evidence in the languages themselves.

The most substantial problems with genetic classification, apart from the aforementioned fact that the relevant evidence tends to disappear with the passage of too much time, involve the fact that innovations are often confusing. There is nothing to stop an innovation from acting on the output of another innovation, sometimes with the result that, by sheer coincidence, it undoes the effect of an earlier innovation. For example, we mentioned in §1.1.2 that Proto-Northwest Germanic changed *[e:] to *[a:], leading to forms like *[la:tan] for 'let' in Northwest Germanic languages. Yet English, a Northwest Germanic language, has [ɛ] in *let*, a vowel that seems much more like a retention of the original Germanic *[e:]. This [ɛ] is the end product of several innovations in English and its near relatives that changed *[a:] to [æ:], then to [ɛ:], then to [ɛ]. The full history makes a coherent story, but innovations on top of innovations are a serious qualification of our earlier assurance that innovations are shared by all descendants and that shared innovations are therefore markers of specific language clades.

A similar problem with innovations is that the same innovation can happen multiple times. For example, the sound change whereby [h] became silent happened in Latin, the ancestor of the Romance languages, leading to silent ⟨h⟩ in words like French *heure* and Spanish *hora*, both from Latin *hora* 'hour'. Subsequently French acquired a new [h] in Germanic loanwords like *hâche* 'ax', and Spanish acquired a new [h] from original [f] in words like *habla* 'talks', from Latin *fabulatur*. Then each of those languages independently introduced an innovation that DELETED those new [h]. So there were at least three separate innovations of deleting [h]. Thus if one sees that a [h] has disappeared, it is not immediately obvious what

clade that innovation is associated with. While that fact may be simply a minor source of confusion for the student learning the history of the Romance languages, repeating innovations can make it hard to correctly reconstruct the cladogenesis of a language family.

Finally, it is important to note that speakers of one language frequently BORROW words or other linguistic structures from another language. Borrowing is just another type of innovation: Latin borrowed *hora* from Greek, then descendants of Latin inherited that innovation. For the person trying to figure out the cladistics of a language family, however, borrowing is a massive annoyance, especially when it occurs between two closely related languages. If two languages have, for example, the same word for 'rice', it could be because they inherited the same word from a common ancestor, or it could be because one or both of the languages borrowed it, perhaps one from the other. That is, having the same word for 'rice' is either a shared innovation, which would suggest that two languages should be grouped together in the same clade, or else it is the result of a borrowing event, which would have no bearing whatsoever on the true cladistic history. Linguists have developed some techniques that often allow them to tell these cases apart, but loans remain difficult all the same. Along with the problem of independent innovations, they are among the foremost reasons that linguists often disagree over the true cladogenesis of a language family.

1.1.4 PROVING GENETIC RELATEDNESS

Genetic classification consists of two broad phases. First, linguists demonstrate that a specific set of languages are related to each other. Second, they may determine the cladogenesis of that set of languages. The traditional, and still by far most common, methodology for performing both of these phases of genetic classification is called the COMPARATIVE METHOD, and it is the topic of the rest of this section.

The comparative method is based on the theory that the relationship between sound and meaning is arbitrary. In principle, if we are presented with words from a language we don't know, we shouldn't be able to deduce what their meanings are, even if we are given a scrambled set of meanings to choose from. If we translated 18 concepts into both Malay and the unrelated language Hausa, we would not expect the Malay translations to have anything in common with their Hausa equivalents. For example, if we looked at Malay words that begin with [t], we would not expect the Hausa equivalents to consistently start with [t] or any other specific phone: we'd expect a random selection of the different sounds that can begin words in Hausa. It follows that if we did the same thing for all Malay and Hausa sounds, counting how often Malay sound X was found at the start of a Malay word when Hausa sound Y was found at the start of the equivalent Hausa word, none of those counts would be particularly high. The same would be true of other positions in the word, not just the initial one. In the terminology of the comparative method,

we would list and count the sound CORRESPONDENCES between the Malay words and the Hausa words. We would end up rejecting the idea that Malay and Hausa were related because of the low number of RECURRENCES.

It should be noted, though, that there are ways to skew these correspondence counts so that several Malay and Hausa words might actually match up, leading to high numbers of recurrences and giving the impression that the languages might be related. One source of problems would be to include several words from that exceptional but not insignificant segment of the vocabulary that is not truly arbitrary. Words that name sounds may sound like those sounds; NURSERY WORDS like those for ‘mother’ or ‘breast’ may sound like the sounds babies all over the world make when they are hungry; words for ‘small’ often contain sharp, low-sonority sounds like [i]. Such words may loosely be grouped under the rubric ONOMATOPOEIA. Loading up our list with concepts like these would increase our chances of finding matches between unrelated languages. Loanwords can also be misleading. Presumably Malay and Hausa have not borrowed any words from each other, but each has borrowed words from world languages like English and Arabic. As we mentioned in §1.1.3, common loanwords could make languages look related to each other when they are not. One solution to these problems is to draw concepts from a list that has been vetted to exclude concepts likely to encourage onomatopoeia or borrowing. Of these, the most widely used is some version of the list in (1), although there have been many proposals for improvements. All such lists concentrate on BASIC VOCABULARY, that is, concepts that are most likely to be expressed by simple, widely known words and are least likely to need borrowing from other languages.

(1) Swadesh 100 list (Swadesh 1955)

I	person	tail	heart
you	fish	feather	liver
we	bird	hair	drink
this	dog	head	eat
that	louse	ear	bite
who	tree	eye	see
what	seed	nose	hear
not	leaf	mouth	know
all	root	tooth	sleep
many	bark	tongue	die
one	skin	claw	kill
two	flesh	foot	swim
big	blood	knee	fly
long	bone	hand	walk
small	grease	belly	come
woman	egg	neck	lie
man	horn	breast	sit

stand	stone	path	hot
give	sand	mountain	cold
say	earth	red	full
sun	cloud	green	new
moon	smoke	yellow	good
star	fire	white	round
water	ash	black	dry
rain	burn	night	name

The number of recurrent correspondences should be much higher when comparing languages that are related to each other (see, for example, Table 1.1). Related languages should have a certain amount of vocabulary retained from their common ancestor. If we compared Malay words to those in another Austronesian language such as Maranao, which is spoken in the Philippines, we would not be surprised to find that words that have Malay [t] also have [t] in Maranao, resulting in a large number of recurrences for the correspondence [t] : [t] (see e.g. the translations for ‘die’, ‘eye’, and ‘louse’). Even if there was a sound change in one or both languages, there would still be some correspondences with a large number of recurrences, because sound change propagates through the whole vocabulary. For example, Hawaiian had a sound change that turned all [t] into [k], so when comparing Malay with Hawaiian, one finds a large number of recurrences for the correspondence [t] : [k].

We will now apply the comparative method to data from real languages spoken in the Pacific (Table 1.1). To keep the example simple, the table contains only a small number of concepts, but otherwise the table is quite illustrative of how linguists might approach a classification study. The data are transcribed in the International Phonetic Alphabet (IPA; see under “Symbols and Abbreviations”) rather than in the ordinary writing systems. The IPA makes it easier to compare the forms, because it frees us from having to learn the orthographies of languages and it eliminates arbitrary features of individual writing systems.

The first step is to formally demonstrate that these languages are all genetically related, by showing that there are an appreciable number of recurrent sound correspondences. Table 1.2 lists all the correspondences that are recurrent, appearing more than once, in the data in Table 1.1. For example, row 3 says that there are two concepts (‘lobster’ and ‘two’) in which [d] in Malay and Maranao corresponds with [l] in Samoan and Hawaiian. Row 9 shows that it is perfectly fine to say that a sound in one language corresponds with nothingness, or ZERO, in another language. Row 13 is a bit more abstract: it takes note of the fact that at the end of a word, Malay and Maranao often agree on having the same consonant where nothing at all is found in that position in Samoan and Hawaiian. With more data, we could have expanded this into a separate correspondence for each of several different consonants.

TABLE 1.1

Vocabulary of four languages of the Pacific

Concept	Malay	Samoan	Maranao	Hawaiian
	MSA	SMO	MRW	HAW
coconut	jiur	niu	niog	niu
cry	taŋis	taŋi	olaʔol	kani
die	mati	mate	mataj	make
drink	minum	inu	inom	inu
eye	mata	mata	mata	maka
five	lima	lima	lima	lima
hardwood	təras	toa	təgas	koa
hibiscus	baru	fau	wago	hau
house	balaj	fale	walaj	hale
I	aku	aʔu	aku	aʔu
lobster	udaŋ	ula	odaŋ	ula
louse	kutu	ʔutu	koto	ʔuku
mistake	salah	sala	salaʔ	hala
pandanus	pandan	fala	ragoroj	hala
path	ḍʒalan	ala	lalan	ala
rafters	kasaw	ʔaso	kasaw	ʔaho
sky	laŋit	laŋi	laŋit	lani
two	dua	lua	doa	lua

TABLE 1.2

Recurrent correspondences from Table 1.1

	MSA	SMO	MRW	HAW	PMP	Concepts
1.	a	a	a	a	*a	cry, ^a die, eye(2×), five, hardwood, hibiscus, house, I, lobster, mistake(2×), pandanus(2×), ^a path(2×), rafters, sky, two
2.	b	f	w	h	*b	hibiscus, house, pandanus ^{a,b}
3.	d	l	d	l	*d	lobster, two
4.	i	i	i	i	*i	coconut, cry, ^a drink, five, sky
5.	k	ʔ	k	ʔ	*k	I, louse, rafters
6.	l	l	l	l	*l	five, house, mistake, pandanus, ^{a,b} path, sky
7.	m	m	m	m	*m	die, eye, five
8.	ŋ	ŋ	ŋ	n	*ŋ	cry, ^a sky
9.	r	∅	g	∅	*R	coconut, hardwood, hibiscus
10.	s	s	s	h	*s	mistake, rafters
11.	t	t	t	k	*t	cry, ^a die, eye, hardwood, louse
12.	u	u	o	u	*u	coconut, drink, hibiscus, I, lobster, louse(2×), two
13.	C _i #	∅	C _i #	∅	*C _i	cry, ^a drink, hardwood, lobster, pandanus, path, sky

^aCognate is lacking in Maranao.^bCognate is lacking in Malay.

From long years of experience, our intuition tells us that we have uncovered more than enough recurrence to demonstrate that the four languages are related, considering that we have only looked at 18 different concepts. But it is impossible to give any specific objective threshold of how much recurrence is enough, because the number will vary depending on such factors as the phonologies of the languages being tested. Statistical tests can give a good estimate of whether the amount of recurrence is significantly higher than what would obtain if the languages were unrelated, but they require a substantial amount of computer modeling (Kessler 2008). In this case, our intuitions are validated by the fact that historical linguists all agree that these four languages are all elements of a large clade known as *Malayo-Polynesian* (§6.1.2), which is a branch of the Austronesian family.

The next step toward subgrouping a set of related languages is to reconstruct what the sounds in each correspondence could have been in the most recent common ancestor. The goal is to minimize the number of sound changes that would be necessary, while maximizing their plausibility and that of the overall inventory of sounds posited for the ancestral language. Obviously, coming up with an optimal solution can be quite complicated, and solutions are often subject to debate, but we give our best guesses in the column labeled PMP (Proto-Malayo-Polynesian) in Table 1.2. The most important consideration is that sound changes must be deterministic. This means that if two different correspondences are to be assigned the same ancestral form, then you must explain what phonetic environment would differ for the different outcomes. This case comes up only once in this analysis, where row 13 indicates that ancestral forms such as [s] are sometimes deleted in Samoan and Hawaiian, while other rows indicate that [s] survives as such or changes to [h]. The cause of these different outcomes is the phonetic environment: deletion happens if and only if the sound is at the end of a word. If the change cannot be predicted from the environment in which it occurred, you must posit different phones in the ancestral language.

1.1.5 SUBGROUPING

The stage is now set for a cladogenetic analysis, which is often referred to as SUBGROUPING. Recall that clades are defined by innovations that are shared by two or more languages. In our data (Table 1.2), Malay and Maranao are very similar to each other, but those similarities are due entirely to retention of the posited Proto-Malayo-Polynesian phone. Therefore the similarities between them are not innovations and must not be taken as evidence of grouping. In contrast, Samoan and Hawaiian share such innovations as the change from *[d] to [l], the change from *[k] to [ʔ], and the deletion of *[R] and of word-final consonants. Furthermore, if we take the Hawaiian change of *[s] to [h] as an instance of a broader change that affects all fricatives (noisy continuant sounds like [s] and [f]), then correspondence 2 can be taken as showing a shared innovation between Samoan and Hawaiian even

though the outcomes are different. If a common ancestor of Samoan and Hawaiian changed Proto-Malayo-Polynesian *[b] to *[f], then the aforementioned rule that changes fricatives to [h], a rule that we need anyway to account for *[s], will turn that intermediate *[f] into the [h] that is attested in Hawaiian. Thus Samoan and Hawaiian share five innovations: deletion of [ʀ], deletion of final consonants, change of [k] to [ʔ], change of [b] to [f], and change of [d] to [l]. In contrast, there are no innovations shared between Malay and Maranao. Therefore Samoan and Hawaiian group together, that is, there was probably an intermediate proto-language between those languages and Proto-Malayo-Polynesian. Accordingly, we have schematized the interrelations among these languages in the tree in Figure 1.3. As far as it goes, this tree agrees with the general scholarly opinion on the relationship between these four languages: Samoan and Hawaiian belong to a group called the *Polynesian* languages, and Malay, Maranao, and the Polynesian languages have no common ancestor later than Proto-Malayo-Polynesian.

The most important lesson to take away from this subgrouping exercise is that languages do not form a clade just because they look particularly similar to each other. Malay and Maranao look alike because both are comparatively similar to their ancestor, not because they have changed in a similar manner. Because of their lack of shared innovations, they do not qualify as being more closely related to each other than to any other Malayo-Polynesian language.

Another reasonable conclusion from this exercise is that genetic analysis of languages can be difficult. Small differences in how the data are interpreted can change the analysis significantly. It is fairly unusual for a language clade of any appreciable size not to be the subject of some controversy between different experts.

The vast majority of languages are susceptible to cladistic analysis. Chapters 3 through 7, the bulk of this book, therefore follow a long linguistic tradition in discussing languages by families, grouping them together with other languages in the same clades. But some languages defy the assumption of the comparative method, that languages develop from a single parent language by gradual change. Mixed languages have two parent languages that are intertwined so thoroughly that assigning them to a specific clade would be undesirably

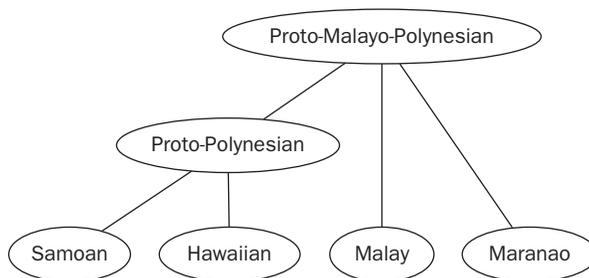


FIGURE 1.3 Cladistic arrangement of four Pacific languages

arbitrary. Constructed languages, pidgins, creoles, and sign languages come into being abruptly, invented by an individual or a group to satisfy some extraordinary communicative need. They are often inspired by one or more existing languages, but the changes involved, especially in the grammar, are typically so thoroughgoing that it would be misleading to characterize them as an ordinary continuation of a single parent. Because such languages cannot be assigned to linguistic clades in the usual sense of the term, they will be discussed in sections of their own, in Chapter 8.

1.2 Typological classification of languages

Languages are also classified by the type of their structure. This is known as **TYPOLGICAL CLASSIFICATION**, or simply **TYPOLGY**. Typologists often classify languages according to the regular linear arrangement of the verb, subject, and object of a clause. Irish and Classical Arabic are verb–subject–object (VSO) languages, because the verb usually precedes the subject, which precedes the object. Hindi and Japanese are subject–object–verb (SOV) languages, reflecting the fact that the verb tends to come at the end of the clause, after the object. Typological classification is valued as a source of insights about language structure – a core concern of modern linguistics. Genetic classification, by contrast, can make only limited claims about language structure. Irish and Arabic have the same VSO word order yet no genetic connection. Irish and Hindi have different word order yet belong to the same family.

Genetic classification aims for a single correct historical analysis. Typological classification is fundamentally different: there is no single correct analysis for any set of languages. Since typology can be based on many different types of linguistic structure, and since structures vary from language to language, even within genetically close groups, it's not uncommon for languages that class together by one typological criterion to class with a different set of languages by other typological criteria. Hindi agrees with Japanese in being SOV, but the two differ greatly in other structural properties: verbal systems, types of modifiers, and phonologies. Still, there are quite a number of correlations among typological features, illustrated later under the rubric *implicational universals*.

Two nineteenth-century German linguists, Friedrich von Schlegel and Wilhelm von Humboldt, introduced the first notions about typology to gain widespread acceptance. They noted significant differences in the ways that languages formed words and showed that the differences could be reduced to a small number of types. A good account of the early history of language classification is Horne (1966). Some of their ideas still hold sway today, though the field of typology has advanced significantly and has expanded to include syntax and phonology as well as morphology. Much recent typological work has, in fact, been multidimensional, combining phonology, morphology, and syntax.

1.2.1 MORPHOLOGICAL TYPOLOGY

Greenberg (1954) has one of the more useful adaptations of Humboldt's morphological typology. Although the traditional typology introduced concepts that are still at the core of how most linguists think about morphological types, the terminology was vague and impressionistic. Greenberg broke the system down into its logical components and proposed objective measures, or INDICES, for each component.

As a first step, we take a large, representative sample of a language, and break each of its words down into MORPHEMES: the shortest strings of phonemes that have some meaning or function in the language. For example, if the text contained the word *illogical* [ɪlɑdʒɪkəl], it would be broken into the parts [ɪ-lɑdʒ-ɪk-əl], because each of those parts can be found with the same function in other words. Greenberg's first index is the average number of morphemes per word. He called this index a measure of SYNTHESIS (from Greek 'putting together'). For English, he reported an index of 1.68 morphemes per word. Although the language has many complex words, like the four-morpheme [ɪ-lɑdʒ-ɪk-əl], it also has many simpler words, which bring down the average. Greenberg looked at other languages, including Vietnamese, for which he reported an index of 1.06, which is close to the absolute minimum of 1.00 morphemes per word. At the other extreme, an Eskimo language he analyzed had 3.72 morphemes per word.

Other typological indices report analogous measures for specific types of morphemes. Morphemes may be ROOTS, which carry core lexical meaning, or AFFIXES – elements like prefixes or suffixes, which are best characterized in terms of their linguistic function. To continue our example with *il-log-ic-al* [ɪ-lɑdʒ-ɪk-əl], the morph [lɑdʒ] is the root, meaning approximately 'reasoning'; the other morphemes are affixes: [ɪ-], which negates, [-ɪk], which forms new nouns, and [-əl], which forms adjectives. It is also possible to have multiple roots in a single word, which is called a COMPOUND WORD. For example, *bird-house* is a compound. Greenberg suggested counting the number of roots per word, as a measure of compounding. He also proposed counting affixes per word, broken down by their function. His INFLECTION index counted INFLECTIONAL affixes, those relevant to the language's grammar, including such things as number (e.g. plural *cats* vs. singular *cat*) and tense (e.g. past *walked* vs. present *walk*). His DERIVATION index counted DERIVATIONAL affixes, those that are used to build new lexemes: words that have different meanings and are not just inflectional variants (e.g. *walker* vs. *walk* and *catty* vs. *cat*). An additional count he suggested was to separately report the average numbers of prefixes and suffixes per word. Greenberg's results showed that Vietnamese has a fairly high compounding index, but its affix indices are all zero. Prefixes were relatively rare in the languages he looked at, with some languages showing no prefixes at all – even Eskimo, which has an unusually large number of morphemes per word. An outlier is Swahili, which has a prefixation index three times as high as its suffixation index.

Lastly, we consider Greenberg's AGGLUTINATION index. This was defined as the proportion of adjacent morpheme pairs that combine in a perfectly regular, predictable fashion. For example, to form the gerund-participle in English, all verbs add *-ing*, with no surprises; all verbs, like *walk-ing* and *be-ing*, add the suffix in an agglutinative fashion. By contrast, the past tense form is often irregular: forms like *slep-t* [slɛp-t] from *sleep* [slip] and *caugh-t* [kɔ-t] from *catch* [kæʃ] show an unpredictable (compare *beeped*, *latched*), non-agglutinative pattern. In Greenberg's study, languages varied dramatically in their agglutination index, ranging from .03 for his Eskimo sample to .67 for Swahili. English was in the middle of the pack, at .30.

Linguists wanting a quick characterization of a language's morphological type bypass Greenberg's careful quantitative technique and typically apply one of the familiar 19th-century terms on an impressionistic basis (2). A language is characterized as ANALYTIC if its synthesis index is low, SYNTHETIC if it is higher, and POLY-SYNTHETIC if it is very high. Not rarely, however, people use those terms to mean something more like Greenberg's inflection index, which ignores compounding and derivational affixes. For example, Middle (Classical) Chinese (LTC), which makes extensive use of compounding but has few inflectional affixes, is often characterized as analytic. The agglutination index is most relevant to synthetic languages, because analytic languages have very few morpheme boundaries and polysynthetic languages are almost all non-agglutinative. Therefore, instead of saying that a language is synthetic, linguists often specify whether it is agglutinative or fusional.

(2) Morphological types

Analytic:	Middle Chinese
Synthetic	
· Agglutinative:	Turkish
· Fusional:	Latin
Polysynthetic:	Yup'ik

A few typical sentences from Middle Chinese, Turkish, Latin, and Yup'ik (3–6) will illustrate the essential differences among the four morphological types. These sentences are presented as interlinear glossed text. This is a common way of presenting data in linguistic publications and will be used throughout this book, so we encourage you to become familiar with its conventions. The top line presents a sentence in the usual, everyday writing system, and the bottom line gives a natural-sounding English translation, inside single quotes. The intervening lines give a detailed analysis, word-by-word and morpheme-by-morpheme. The second line transcribes the sentence in phonemic IPA. As illustrated in (3b), morphemes within a word are separated by hyphens. The next line then explains the text. Note how this line is always aligned with the preceding one on a word-by-word basis. Within a word, morphemes are not aligned as such, but you can easily see which explanation goes with which morpheme, because there is one explanation

per morpheme, with matching hyphens. Generally speaking, root morphemes are expressed by an equivalent English word. Morphemes that express a more functional meaning, such as affixes, are described rather than translated. Description is done by writing standard abbreviations in small capital letters. Abbreviations will be defined where introduced for the first time, but see also the complete list of abbreviations at the front of this book. The *Leipzig glossing rules* (Max Planck Institute 2008) explain the most important glossing conventions.

(3) Analytic type: Classical Chinese

- a. 下馬入車中。Xià mǎ rù chē zhōng.

xja\ ma\ ɹu\ tʃʰə\ tʃuŋ\
descend horse enter chariot middle

‘[He] got off the horse [and] got into the chariot.’

- b. 松下問童子。Sōng xià wèn tóngzǐ.

suŋ\ xja\ wən\ tʰuŋʅ-tsi\
pine under ask lad-DIM

‘Under the pine trees [I] asked the boy.’

DIM = DIMINUTIVE FORM, which expresses smallness.

- c. 豬人立而啼。Zhū rén lì ér tí.

tʃu\ ɹən\ li\ ɹʅ\ tʰi\
pig person stand and cry

‘The pig stood up [like a] person and cried.’

- d. 楚莊王賜群臣酒。Chǔ Zhuāng wáng cì qún chén jiǔ.

tʃʰu\ tʃwaŋ\ waŋ\ tsʰi\ kʰyn\ tʃʰən\ kjəu\
Chǔ Zhuāng king bestow group minister wine

‘King Zhuāng of Chǔ bestowed wine on his ministers.’

The Classical Chinese text is here cited in traditional Chinese characters and also in an official Chinese spelling in Latin letters, PĪNYĪN [pʰin˥lin˥]. Both the pinyin and the IPA given here reference the modern Mandarin pronunciation (§4.10.2).

The morpheme [xja\] can act as a verb, as in sentence (3a), where it was translated as ‘descend’, or as a POSTPOSITION, as in sentence (3b), where it is translated as ‘under’. A postposition is like a preposition, except that it is placed after the noun phrase instead of before it. Likewise [ɹən\], which usually functions as a noun ‘person’, appears in an adverbial function in sentence (3c) without any formal marking that it is an adverb. This illustrates one of the secondary traits of analytic languages: the same morpheme may sometimes act as a different part of speech depending on the context, without any morphological marking to signal the different function. Sentence (3d) illustrates the fact that word order alone marks

three different grammatical relations among the noun phrases: the subject [tʰuɔ̃ wɑ̃ŋ] 'King Zhuāng', the indirect object [kʰynʰ tʰənʰ] 'his ministers', and the direct object [kjəu] 'wine'.

There are no true affixes in these examples; all the morphemes cited are roots. The only morpheme that seems to act as an affix is [tsi], the diminutive (3b). However, this morpheme appears also as an independent word meaning 'son' or 'offspring'; therefore [tʰuŋʰ-tsi] 'lad' is best analyzed as a compound.

No morphemes cited in the previous examples have any ALLOMORPHS, insofar as this can be determined from the logographic writing system. By *allomorphs* we mean pronunciation variants of the same morpheme, as in the difference between English *child* [tʃaɪld] and *child-ren* [tʃɪld-rɪn].

(4) Agglutinative type: Turkish

- a. Köyünden çıkmamış köylü bu meseleleri anlarmı?
 köj-yn-'den tʃʷk-ma-muʃ köj-'ly
 village-3.POSS-ABL come-NEG-PST.PTCP village-inhabitant
- bu mesele-le'r-i an'la-r-mu
 DEM.PROX problem-PL-ACC understand-HAB-Q

'Does the villager who has not left his village understand these problems?'

3 = THIRD PERSON, referring to an entity that does not include the speaker or the addressee.

POSS = POSSESSIVE, marking possession or some other dependency relation between nouns.

ABL = ABLATIVE CASE, denoting starting point of movement.

NEG = NEGATIVE POLARITY, 'not'.

PST = PAST TENSE.

PTCP = PARTICIPLE, a verb form used like an adjective or adverb.

DEM = DEMONSTRATIVE, a word class whose function is to point out what item the speaker is referring to, such as English *that* and *this*.

PROX = PROXIMAL DEMONSTRATIVE, something nearby.

PL = PLURAL NUMBER, referencing more than one instance of something.

ACC = ACCUSATIVE CASE, marking the direct object of a verb (here, [an'larmu] 'understand'; see §3.5.4.1 for more about case).

HAB = HABITUAL ASPECT, marking an action that occurs frequently (see §3.5.4.5 for more about aspect).

Q = INTERROGATIVE, marking a question.

Dots are used in the explanation line to separate concepts that are just one morpheme in the language being described. Thus PST.PTCP means that [-muʃ] marks a past participle, and DEM.PROX means that [bu] is a proximal demonstrative, 'this'.

b. Evlerimizden gelmiyordum.

- ev-ler-im-iz-'den 'gel-mi-jor-d-um
 house-PL-1SG.POSS-PL.POSS-ABL come-NEG-PROG-PST-1SG

'I was not coming from our houses.'

1 = FIRST PERSON, referring to the speaker or a group including the speaker.

SG = SINGULAR NUMBER, referring to one person or thing.

PROG = PROGRESSIVE ASPECT, marking a continuous, ongoing action or state.

It is evident from these examples that both nouns and verbs in Turkish may consist of relatively long strings of morphemes: a root followed by a series of suffixes. What is not readily apparent from the examples is that allomorphy is overwhelmingly regular. In addition there is a marked absence of portmanteau morphemes; concepts such as number, case, tense, aspect, and person are all marked by separate, independent morphemes.

(5) Fusional type: Latin

a. Filius patrem amat.

fi:lɪ-ʊs patr-ɛm am-at
 offspring-M.NOM.SG father-NN.ACC.SG love-3SG.PRS.IND

‘The son loves the father.’

M = MASCULINE GENDER, a class of nouns that includes words referring to men (see §3.5.4.1 for more about gender).

NOM = NOMINATIVE CASE, which marks a noun as being the subject of a verb.

NN = NON-NEUTER GENDER, pertaining to any gender in the language other than neuter, i.e. in Latin either masculine or feminine (the N- means ‘not’; N by itself means ‘neuter’).

PRS = PRESENT TENSE.

IND = INDICATIVE MOOD, which marks a verb as expressing a fact.

b. Pater filium videt.

pater fi:lɪ-ʊm vid-et
 father[NOM.SG] offspring-M.ACC.SG see-3SG.PRS.IND

‘The father sees the son.’

Square brackets surround grammatical categories that are expressed by the absence of an affix in a context that otherwise has such affixes. Unlike [fi:lɪʊs] in (5a), [pater] has no nominative singular suffix. In some theories, such a lack of a morpheme is referred to as a *zero* morpheme.

The examples show only a moderate amount of affixation, which is normal for fusional languages. There are several examples of portmanteau morphemes, such as [-ʊs], which simultaneously expresses masculine gender, nominative case, and singular number. Moreover, it is not always clear where the morphological cuts should be made. Consider, for example, the suffixes that are here glossed as markers of the third person singular present indicative: [-at] and [-et]. Some verbs take the suffix [-at], and others take the suffix [-et], and yet others take [-it], which is not illustrated here. The way we glossed these suffixes in the examples suggests they are essentially allomorphs, opaquely conditioned by the verb. Alternatively, the fact that all third person singular verbs end in [-t] suggests that [-t] should be isolated

as the marker of the third person singular. That would mean that the preceding vowels, such as [-a-], would have to be analyzed some other way. Perhaps [-a-] could be considered part of the verb itself ([ama-t]), or a morpheme that marks the present indicative, or the present tense in general; or perhaps it is a semantically empty morpheme that simply marks the conjugational pattern of the verb. All of these analyses have been seriously proposed, and it is very difficult to choose between them. This difficulty is at the heart of the definition of a fusional language.

Finally, it is typical of fusional (as well as polysynthetic) languages that most root or stem morphemes are BOUND. That is, they cannot appear alone as independent words in the language. This is true of most of the roots given in the two preceding Latin examples.

Central Alaskan Yup'ik, an Eskimo language, furnishes examples of a polysynthetic language.

(6) Polysynthetic type: Yup'ik

a. Angyarpaliyugngayugnarquqllu.

an̄jaβ-pa-li-juγŋa-juγnaβ-quq=̄tu
boat-big-make-be_able-probably-INTR.3SG.IND=also

'He can probably make big boats, too.'

INTR = INTRANSITIVE VERB, one that takes no direct object.

The low line “_” in “be_able” stands in for a space when multiple words are needed in an English gloss. An actual space might mislead the reader into thinking that “be” and “able” gloss two separate Yup'ik words.

The double-hyphen “=” connects CLITICS to their host. A clitic is a morpheme that behaves in many respects like a free word, but forms a phonological unit with some other word.

b. Angutem neraa neqa.

aŋut-ə̄m nə̄β-a: nə̄qa
man-ERG.SG eat-TR.3SG.IND fish

'The man eats the fish.'

ERG = ERGATIVE CASE, which marks the subject of transitive verbs.

TR = TRANSITIVE VERB, i.e. one that takes a direct object (here, [nə̄qa] 'fish').

In languages of this type, word order is not as important as the order of morphemes in a word. Polysynthetic languages often express in one word what would in other languages be a verb and its arguments, such as a direct object: note how example (6a) has morphemes for 'make' and 'boat' combined in the same word. In Yup'ik this combination of verb and object is somewhat constrained: there aren't too many different verbs that enter into such constructions, and the nouns typically act in an indefinite, generic sense; compare (6b), where a definite object is expressed as a separate word. The combination of a verb and its argument in one word has been called NOUN INCORPORATION, although some linguists prefer to restrict that term to more complicated situations where the syntax permits verbs to absorb their objects.

Besides having portmanteau morphemes, such as the suffix in (6b) that simultaneously expresses transitive, third person, singular, and indicative, Yup'ik has extremely complicated MORPHOPHONEMICS – phonological rules that apply differentially to different morphemes. Such rules render many a morpheme boundary opaque. For example, some suffixes cause the preceding consonant to be deleted, but others do not. As in other polysynthetic languages, morphemes used in noun incorporation may look quite different from the way they look as independent nouns and verbs.

At this point the reader may wonder which of the above four types of languages English belongs to. The answer to this question is not, unfortunately, a straightforward one. Although English shares many features with analytic languages such as Middle Chinese, it also shares some features with fusional languages such as Latin. Arguing for an analytic designation is the fact that, at least in the basic, native vocabulary, English words are short. English often dispenses with affixes that would be required in many other languages. For example, the English words *fish* and *man* can both be used as either verbs or nouns without any derivational affix to indicate the difference. Most roots and stems can appear as a free word in a sentence, and English makes free use of compounding. In English, too, it is the word order alone that marks the grammatical relations among the three noun phrases in sentences like *Peter gave John the book*.

On the other hand, English has a lot of irregular morphology, including SUPPLETION (basing different inflected forms on different roots, like *go* vs. *went*) and portmanteau allomorphs (e.g. *feet* is a form of *foot* that simultaneously expresses the feature of plurality). These features make English look more like a fusional language. In addition, especially in Latinate and scientific vocabulary, there are words containing long sequences of affixes, suggesting the type of structures common in synthetic languages (e.g. *anti-dis-establish-ment-ari-an-ism*). Thus, one may conclude that, typologically speaking, English is neither fish nor fowl: it is a typological anomaly.

It turns out that most of the world's languages do not fit neatly within one of the four categories of traditional morphological typology, whereas all can be precisely characterized in terms of Greenberg's indices. Nevertheless it is much more common for people to report that a language is analytic, fusional, agglutinative, or polysynthetic rather than to give a series of indices for a language. These terms give only a general impression of the nature of a language's morphology, but they are much more convenient to report and also much more memorable for the reader than a list of numbers.

1.2.2 SYNTACTIC TYPOLOGY

The most common way of characterizing the syntax of a language is to report its WORD ORDER. The convention is to report the relative order of a transitive verb (V)

and its subject (S) and object (O) nouns. Thus if the subject precedes the verb and the object follows it, the word order is SVO; VSO would mean that the verb comes before the subject, which comes before the object. (7) illustrates SOV order; its English translation is SVO:

(7) Japanese

táloo ga d̥z̥íloo o m'imáçita
 Tarō NOM Jirō ACC saw
 S O V

'Tarō saw Jirō.' (Martin 1962: 66)

MORPHOSYNTACTIC ALIGNMENT tells how case markings on nouns correspond with the grammatical relations of subject and object. In NOMINATIVE-ACCUSATIVE languages, the subject of a transitive verb and the subject of an intransitive verb have the same case marking, which is usually called the NOMINATIVE CASE (NOM). In contrast, a direct object of a transitive verb is marked as having a different case from the subject, usually called the ACCUSATIVE CASE (ACC).

(8) Latin

a. Paulus Homerum legit.

paul-us hōme:r-um lēgit
 Paul-NOM Homer-ACC reads
 SUBJECT OBJECT V_{TR}

'Paulus is reading Homer.'

b. Homerus dormitat.

hōme:r-us dōrmitat
 Homer-NOM dozes
 SUBJECT V_{INTR}

'Homer nods.'

Thus, in nominative-accusative languages all subjects are marked for the same case, and the contrast is between subjects and objects.

In ERGATIVE-ABSOLUTIVE languages, however, the subject of an intransitive verb and the direct object of a transitive verb have the same case marking, usually called ABSOLUTIVE (ABS) (9a), whereas the subject of a transitive verb is marked by a different case, called the ERGATIVE (ERG) case (9b). However, as is the case with most universals, these generalizations are not without exceptions, and most ergative-absolutive languages have at least some nominative-accusative constructions (§3.3.3, §3.4).

(9) Basque

a. Gizona kalean dago.

gison-a kale-an dago
 man-ABS street-LOC is
 SUBJECT V_{INTR}

'The man is in the street.'

LOC = LOCATIVE CASE, showing location.

b. Emakumeak gizona ikusten du.

emakume-ak gison-a ikusten du
 woman-ERG man-ABS see AUX
 SUBJECT OBJECT V_{TR}

'The woman sees the man.' (Both examples from Saltarelli 1988: vi–vii.)

Greenberg's interest in language typology led him to take up Roman Jakobson's theory of IMPLICATIONAL UNIVERSALS (1958) and extend it to syntax. In 1966, Greenberg demonstrated that the presence of a certain syntactic feature often implies the presence of one or more other features. For example, he showed that languages that have VSO word order almost always place prepositions before noun phrases and that the majority of SOV languages he looked at were POSTPOSITIONAL, placing equivalent words after noun phrases. This finding can be cited as an implicational universal: if a language has VSO as the basic word order, then it has prepositions instead of postpositions; and if a language has SOV as the basic word order, then it probably has postpositions instead of prepositions. Another implicational universal concerns inflected auxiliary verbs: they precede the main verb in VSO languages, but follow it in SOV languages.

The discovery of implicational universals was a very important development for linguistic typology for two reasons. First, it allows for a more economical typological schema: if languages are described on the basis of those features that entail a large number of other important features, one needs only state that a given language has this basic feature; that will automatically imply a whole series of other features as well. For example, by classifying a language as SOV, it is not necessary to explicitly state that it is also a postpositional language. Such implicational relationships are not limited to syntax but can be found in other components of language as well. An open question is the degree to which important implications span the various components of grammar, so that one could, for example, predict word order on the basis of a language's phonology.

Second, the existence of implicational universals alerts us to many phenomena that require explanation; that is, linguistic theory must explain why there are such interrelationships between various grammatical features. In turn,

our search for such explanations usually leads us to discover even more hidden facts and relationships in language. Thus, what starts as a typological, classificatory endeavor eventually leads to ever expanding understanding of the workings of language.

1.2.3 PHONOLOGICAL TYPOLOGY

In line with typologies based on syntax and on morphology, the task of PHONOLOGICAL TYPOLOGY is to discover worthwhile correlations involving the distribution of sounds. In a pioneering 1963 study, Charles Ferguson listed a number of statements about nasal sounds that appeared to hold universally. For example, no language has only vowels that are nasalized, while some languages have only non-nasalized vowels. We can base an elementary typology on this simple observation, classing languages into one of two types: those with nasalized vowels and those without. We can then examine languages of both types for further generalizations. In a language with nasalized vowels, are some vowels more likely than others to bear nasality? For example, if a language has just one nasalized vowel, will it be low, mid, or high? The answer appears to be that [ã] is favored but not without exception. In languages lacking nasalized vowels, is there a favored vowel? The answer appears to be that [a] is favored.

We also deal with phonological typology when we classify systems characterized by linguistic TONE or STRESS. Tone refers to pitch levels or pitch shapes that can distinguish words. Languages employing tone this way are common in many parts of the world, including Asia (e.g. Mandarin Chinese and Thai), Africa (e.g. Yoruba and Zulu), and to a lesser extent the Americas (e.g. Navajo and some of the Uto-Aztecan languages).

STRESS, on the other hand, is a combination of phonetic features like loudness, pitch, and length that make one syllable prominent in relation to its neighbors. In languages like Russian, the syllable that takes the stress cannot ordinarily be predicted from the segmental phonology of the word; thus the position of stress may distinguish one word from another. In other languages, the position of stress is predictable, as in Czech, which always stresses the first syllable of the word, and Polish, which always stresses the second to last syllable. In English, stress is often but not always predictable.

Some languages have both tone and stress. In Norwegian and Swedish, a word can have one of two tonal CONTOURS (distinctive changes in pitch while pronouncing the tone), and the contour's realization depends on where stress is located in the word. In certain Balto-Slavic languages – Serbo-Croatian and Lithuanian are examples – word stress and tone are independent to an extent, though the two systems interact.

Japanese opens up a new possibility, because it has ACCENT but no stress. Tone rather than stress marks accentual prominence in words. Traditionally this seeming hybrid was labeled a PITCH-ACCENT language. The same designation

has also been used for other systems mixing stress with tone, including those of Norwegian and Swedish. But, as Hyman (2009) has observed, no single set of characteristics unifies these supposed pitch accent languages against putative tone languages and stress languages. Rather than classify languages at all, Hyman's prosodic typology classifies properties – tone and stress – and ways that these properties interact.

1.3 Exercises

1.3.1 ESTABLISHING GENETIC RELATIONSHIP

Table 1.3 contains lexical items from Finnish and Hungarian. Your task is to decide whether these two languages are genetically related, based solely on the given data. If you decide that it is possible to establish the genetic relationship in this case, cite the relevant evidence for your claim. If you decide that it is not possible to do so, discuss the difficulties and your reasoning. Is there any counterevidence or irrelevant data you had to discard? Hint: You may find it easier to concentrate on the beginnings of words, at least as a first step.

1.3.2 MORPHOLOGICAL TYPOLOGY

Examine the following data from four different languages and determine which morphological type is exemplified by each. Justify your decisions.

- (1) təx^w-p-ɛlits'ɛʔ-ən-tʃ-ɛʃ

buy-INCHO-clothing-TR-1SG.OBJ-3SG.SBJ

'He bought me some clothing.'

INCHO = INCHOATIVE ASPECT, which emphasizes the beginning of an activity.

SBJ = SUBJECT and OBJ = OBJECT. 1SG.OBJ means that the (possibly implied) object of the verb is first person singular ('me'), and 3SG.SBJ means that its subject (possibly implied) is third person singular ('he' or 'she').

- (2) t'ika-j-ku-man-ɬa p^hawa-sa-n-ku-tʃu

flower-my-PL-towards-only fly-PROG-PRS-PL-Q

'Are they only flying toward my flowers?'

- (3) ua alu le teine i le fale-maʔi

PFV go DEF.ART girl to DEF.ART house-sick

'The girl has gone to the hospital.'

Both [fale] and [maʔi] are roots that can be used as independent words.

DEF = definite, ART = ARTICLE.

PFV = PERFECTIVE ASPECT, which views an action as a complete whole.

TABLE 1.3
Data for Exercise 1

	Gloss	Finnish	Hungarian
1.	child	lapsi	ɟermek
2.	church	kirk:ɔ	tɛmplom
3.	six	ku:si	hɔt
4.	fire	tuli	ty:z
5.	head	pæ:	fɔ:
6.	three	kolme	ha:rom
7.	cookie	keksi	keks
8.	water	vesi	viz
9.	fish	kala	hɔl
10.	sun	aurinko	nɔp
11.	winter	talvi	te:l
12.	sugar	sokeri	ʈsukor
13.	tree	pu:	fɔ
14.	tongue	kieli	ɟɛlv
15.	eye	silmä	sɛm
16.	blood	veri	ve:r
17.	death	kuolema	hɔlɑ:l
18.	zero	nol:a	nul:ɔ
19.	son	poika	fiu:
20.	under	al:a	ɔlɑ:
21.	hear	ku:l:a	hɔl:
22.	bee	mehiläinen	me:h
23.	nose	nenä	or:
24.	soap	saip:ua	sap:ɔn
25.	cheep	pi:p	ʈʃip
26.	give	anta:	ɔd
27.	nest	pesä	fe:sek
28.	meow	miau	miɔu
29.	breast	rinta	mɛl:
30.	half	puolikas	fe:l
31.	go	men:ä	mɛn:i
32.	know	tuntea	tud
33.	heart	sydæn	si:v
34.	butter	voi	vɔj
35.	hundred	sata	saz
36.	mouth	su:	saj
37.	what	mitä	mi

- (4) j-əs k^hol-ə mat^hi sǎg^h-u t^hi-jo
that-PREP river-PREP across bridge-M.NOM was-3SG.PST.IND

'Across that river there was a bridge.'

PREP = PREPOSITIONAL CASE, used for the object of a postposition.

1.3.3 QUANTIFICATIONAL TYPOLOGY

Choose two different passages of English of about 200 words each and perform a simple morphological analysis on them. One of the passages chosen should consist of a running conversation, whereas the other should be an example of a narrative in formal, literary style. Break the words down into morphemes and calculate the degree of synthesis and the compounding index for the two passages. If either of the indices is significantly different for the two passages, how do you account for the difference?

1.4 Suggested readings

Full bibliographic details for each item can be found in the References section in the back of the book.

1.4.1 ONLINE REFERENCES

- ✎ *Ethnologue* (Lewis et al. 2016). Includes an index of 41,186 names for 7,413 current languages and a bibliography with 25,000 citations. Regular updating helps keep the classification current. <http://www.ethnologue.com>
- ✎ Glottolog 2.7 (Hammarström et al. 2016). Listing of the world's families and languages in each of their varieties, with their classification and extensive bibliographical references. <http://glottolog.org>
- ✎ MultiTree: A digital library of language relationships (MultiTree 2014). Interactively displays related languages in a hierarchical tree, with constantly updated content. <http://multitree.org>
- ✎ The world atlas of linguistic structures online (Dryer & Haspelmath 2013). Lists languages that have specified structural features and plots their location on a map of the world. <http://wals.info>

1.4.2 SURVEY ARTICLE

- ✎ Languages of the world (Comrie 2001). An overview of the world's languages in just over 20 pages.

1.4.3 ENCYCLOPEDIAS AND COLLECTIONS

The collections below offer an overview of different language families, often sketching linguistic structure in some of their languages. The entries in these works, all by respected specialists, differ in format, in the number of languages covered, and in the amount of space devoted to each language or family.

- ✎ *The atlas of languages: The origin and development of languages throughout the world* (Comrie et al. 2003). For nonlinguists, a survey

of languages of the world with a wealth of color maps and pictures of their speakers and locations.

- ✘ *Encyclopedia of language and linguistics* (K. Brown 2008). This massive compendium has sections on the world's major language families and on member languages.
- ✘ *One thousand languages: Living, endangered, and lost* (Austin 2008). Covers the languages of the world by region, including endangered and extinct languages.
- ✘ *The Routledge concise compendium of the world's languages* (G. Campbell & King 2011). Over 100 chapters, each describing the linguistic structure of a major world language and its cultural and historical background.
- ✘ *The world's major languages* (Comrie 2009). Chapters on the linguistic structure of 50 major world languages, each by a specialist in that language.

1.4.4 GENETIC CLASSIFICATION

- ✘ *Historical linguistics: An introduction* (L. Campbell 2013). Popular basic textbook using examples from around the world, but principally from Europe and North America.
- ✘ *Language classification: History and method* (L. Campbell & Poser 2008). A comprehensive, scholarly treatment of past and current controversies in language classification, with detailed critiques of methods and of a number of genetic groupings, some of them widely accepted.
- ✘ *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (Hock & Joseph 2009). Introductory textbook, with most of the examples drawn from Indo-European languages.
- ✘ Nostratic (Kaiser & Shevoroshkin 1988). An overview of the Nostratic hypothesis written by its proponents.
- ✘ *Principles of historical linguistics* (Hock 1991). A thorough, well-written advanced-level introduction to historical linguistics and the comparative method. Along with excellent books that are more recent, this one is still worth consulting.
- ✘ *Trask's historical linguistics* (Millar 2015). Introductory textbook, with a case study at the end of each chapter.

1.4.5 QUANTITATIVE METHODS

- ✘ *Language classification by numbers* (McMahon & McMahon 2005). Introduction to computational methods for language classification, including lexicostatistics.
- ✘ Preliminary lexicostatistics as a basis for language classification: A new approach (G. Starostin 2010).

- ❏ *Quantitative methods in linguistics* (K. Johnson 2008). Chapter 6 has an informative and richly illustrated introduction to computational methods for classifying languages, with practical guides to software use.
- ❏ Tutorial on computational linguistic phylogeny (Nichols & Warnow 2008). A 60-page survey of computational methods in historical linguistics.

1.4.6 TYPOLOGICAL CLASSIFICATION

- ❏ *Introduction to typology: The unity and diversity of language* (Whaley 1997). Broad introduction to the different areas of linguistic typology that assumes little or no background in linguistics.
- ❏ *Language typology and language universals: An international handbook* (Haspelmath et al. 2001). Comprehensive survey of typological phenomena and theory, with over 100 articles.
- ❏ *Language universals and linguistic typology: Syntax and morphology* (Comrie 1989). An overview of developments in language typology; very readable, though not an elementary introduction to the topic.
- ❏ *Linguistic diversity in space and time* (Nichols 1992).
- ❏ *Linguistic typology: Morphology and syntax* (Song 2001). Introductory textbook with chapters on the major areas of syntactic typology.
- ❏ *The Oxford handbook of linguistic typology* (Song 2011). 30 chapters and over 650 pages on typology and its subareas, especially as typology relates to linguistic theory and methodology.
- ❏ *Typology and universals* (Croft 2002). This work's emphasis and general orientation make it an excellent companion volume to Comrie (1989).
- ❏ What, if anything, is typology? (Nichols 2007).

Classification of Writing Systems

Writing is a means of communicating graphically. Not every language has a written form. It appears that writing has only existed for the past 5,000 years, a fraction of the time language has existed; indeed, only in the past few centuries has writing spread to more than a handful of languages. Where writing does exist, it is essentially a means of representing speech. These considerations have led linguists to classify writing as a secondary concomitant of language.

Nevertheless, there are many reasons to study writing. As a tool that people have developed to represent language, writing provides insights as to how humans conceptualize language. Writing is often an important route by which linguists get access to language samples, making it important for linguists to know how writing works and to discover its strengths and limitations. Perhaps most important, it is vital for linguists to understand ways in which the different structures of writing systems do and do not correspond to differences in the structures of the languages they represent.

Just like languages themselves, writing systems can be classified either genetically, according to their historical origin, or typologically, according to various criteria. We will discuss their typological classification first.

2.1 Typological classification of writing systems

Writing systems differ from each other in many ways; it is therefore necessary to choose which criteria to base the typological classification on. For example, one may choose to classify writing systems according to the shapes that they employ for their symbols. Although this is not a very enlightening type of classification, it would have some practical use: it could help us design a writing-system identification guide. Writing systems can also be classified by the DIRECTION OF WRITING. All writing is SEQUENTIAL, with one symbol, or CHARACTER, following another, but the direction varies. Horizontal, left-to-right writing is the most common, but several scripts, including Arabic and Hebrew, are written from right to left, and a

few, notably the classical Chinese script, are written vertically. No one has found any linguistic significance to directionality.

Linguists are usually more interested in classifying writing systems by how they express meaning. We begin our classification by defining WRITING as a system of communication that uses a set of graphic symbols and a grammar for combining them to express an unlimited number of messages. Let's begin exploring this definition by looking first at important types of communication that fulfil part but not all of its terms.

GRAPHIC COMMUNICATION consists of making marks on some object in order to convey a message. Historically important means of making marks have included scratching and carving – the word *carve* is COGNATE to (has a common ancestor with) the Greek root of *graphic*, *γραφ* – as well as painting, applying ink to paper with a pen, and stabbing wet clay with a reed. Nowadays, of course, the use of any modern technology that produces an analogous visual effect is also considered graphic. Table 2.1 uses indentation to hierarchically outline the types of graphic communication in terms of how they convey meaning. For example, it says that syllabographic and phonemographic symbols are types of phonographic symbols, which are types of glottographic symbols, and so forth; these concepts will all be discussed in this chapter.

Writing is a type of symbolic communication and not a type of iconic communication. The words ICONIC and SYMBOLIC are drawn from Peirce's classification of communicative signs (Peirce 1984: 56). In Peirce's analysis, SIGNS are relations between forms and the meanings they take on in the mind of an interpreter. He used the term *icon* to designate forms that communicate only by dint of the fact

TABLE 2.1
Types of graphic signs

Type	Section
iconic	§2.1
symbolic	
· nonsystematic	
· writing	
· · semasiographic	
· · glottographic	§2.1.1
· · · logographic	§2.1.2
· · · phonographic	§2.1.3
· · · · syllabographic	§2.1.3.2
· · · · phonemographic	§2.1.3
· · · · · alphabet	§2.1.3.1
· · · · · syllabically organized	
· · · · · · abjad	§2.1.3.3
· · · · · · abugida	§2.1.3.4

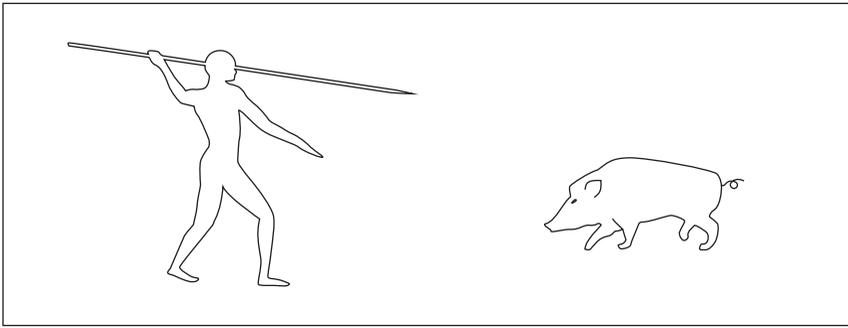


FIGURE 2.1 Iconic representation. Credit: Daniel Leben-Wolf.



FIGURE 2.2 Graphic symbols: Biohazards, poison, and air transportation. Credits: biohazard symbol, Sisor and Nerd65536 on WikiMedia Commons; poison, Daniel Leben-Wolf; air transportation symbol, AIGA.

that they resemble the thing they are meant to represent. Figure 2.1 is our attempt to iconically communicate the message that a man is spearing a pig. Iconic representation is like writing in that it is a (quasi)permanent form of graphic communication, but people tend to classify it as something closer to art than to writing: we *drew* Figure 2.1, we didn't *write* it.

Peirce contrasted icons with symbols, which communicate because the sender and interpreter of some shape have already agreed on its meaning or learned some previously established convention. For example, the first sign in Figure 2.2 is the standard symbol warning of biohazards. It is a symbol because it conveys a specific piece of information – you need to take precautions against exposure to biological hazards here – and it does so because people have learned that that's what it means. Indeed, Charles Baldwin and the Dow Chemical company intentionally designed it to be as abstract and nonrepresentational as possible, so that lab managers would understand that everybody would have to be educated as to its meaning and implications for health and safety (J. Cook 2001).

It is easy to classify a sign as a symbol when it doesn't look like anything. But graphic symbols are often pictorial, at least in origin. What makes them different from icons is that people learn that others have attached a particular meaning to a

symbol. For example, the skull and crossbones in Figure 2.2 is a conventional symbol for poisons. If it were just an icon, its direct meaning would be ‘a human skull and a couple of humeri’, and it would be up to the reader to guess why somebody chose to draw those bones on a bottle of tasty-looking liquid. Also, the next artist who wanted to warn us about a poison might draw a different collection of bones, or come up with some quite different way to express the concept of poison.

Graphic symbols play an important role in modern life because they can be more compact than a verbal message and can be used equally well by people who speak different languages. The user interfaces to smart phones rely heavily on such small symbols. To take another example: In the 1970s, the AIGA and the U.S. Department of Transportation developed a system of 50 graphic symbols for directing people to sites likely to be of interest to visitors. Now, whenever someone wishes to put up a sign directing people to an airport, one can use the airplane symbol in Figure 2.2, without having to decide how to picture an airport or convey that information in all the languages tourists speak.

Sets of symbols can be quite large. But what can really expand the communicative range of symbol sets is not so much the number of symbols, but a GRAMMAR for combining existing symbols to express additional concepts. Mathematical symbols are an excellent example. In the familiar Hindu-Arabic system, whole numbers are built up from only 10 symbols: 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. By themselves, these symbols let people express 10 different numbers. By adding just one grammar rule to the system, power-of-ten place notation, a potentially infinite range of whole numbers can be written, such as 42 or 999. Other important rule-based symbol systems are musical notation and chemical formulas. These are all considered types of *writing*: people *write* numbers and sheet music. They fulfil our definition of being systems of communication that use sets of graphic symbols and grammars for combining them to express an unlimited number of messages.

2.1.1.1 GLOTTOGRAPHY

Powerful as symbolic writing systems are, most of them can be used only in a particular domain. It would be very hard to invent graphic symbols and rules for communicating about all noteworthy concepts in human experience. A few philosophers have claimed that they did so, but because their inventions have never been put into practice, it is impossible to demonstrate that they are really adequate for general communication. For a truly general graphic communication system, only one solution has ever proved practical: inventing symbols not for arbitrary concepts, but for units of language. After all, people are largely happy with the range of topics they can discuss in speech. If a writing system can represent language, then it indirectly can express the wide range of ideas one can express with language. A writing system that represents language is called a GLOTTOGRAPHY, from a Greek root for ‘tongue’ (γλωττ-); in contrast, a writing system like musical notation that doesn’t rely on language to represent

ideas is called a SEMASIOGRAPHY (Greek σημασι- ‘meaning’; W. Haas 1976). Very often those technical terms are dispensed with, especially by those who argue that only glottography is true writing and that semasiography is at best a primitive precursor of writing. While we, as linguists, do particularly admire glottography, we note that if one delves deeply enough into writing systems and their historical development, semasiographic elements are present, and it is important to be able to recognize them.

Glottographic writing is subclassified according to the kind of linguistic units the basic symbols of the writing system represent. If one hopes to develop a set of symbols that is both small enough to be learnable yet expressive enough to write any possible sentence, there are only a few reasonable choices. A natural idea is to develop one symbol for each word in the language: a LOGOGRAPHIC approach (Greek λογ- ‘word’). This is a sensible choice, because words are more or less finite in number, have accessible meaning, and function as the basic elements of syntax. We examine this possibility first, in section 2.1.2. Another idea is to develop one symbol for each sound in the language: a PHONOGRAPHIC approach (Greek φων- ‘speech sound’), the topic of section 2.1.3. It is hard to draw a rigid theoretical boundary between these two types of writing, because, after all, one aspect of words is their sounds. The main difference is that the logographic approach aims to have a unique symbol for each word, without any specific effort to represent its phonology analytically, in a consistent way across all words.

2.1.2 LOGOGRAPHY

A LOGOGRAM is traditionally defined as a character that represents a word. The element *gram*, Greek γραμμ-, refers to a unit of graphic communication; it is derived from *graph*. Because logograms in practice also represent the individual roots of a compound word and sometimes prefixes and suffixes, a more precise definition is that a logogram represents a morpheme (review §1.2.1 for more about morphemes). A morpheme has a pronunciation and a meaning or function. For example, *in-aud-ible* is composed of three morphemes: [ɪn] ‘not’, [ɔd] ‘hear’, and [əbl], a suffix that turns verbs into adjectives. English also has a prefix [ɪn] meaning ‘in’ or ‘into’, as in the phonetic term *ingressive* ‘(air) going into (the body)’. Because its meaning is totally different from that of [ɪn] ‘not’, those two are different morphemes. Conversely, because *hear* has a totally different pronunciation from *aud*, the two are different morphemes, even though the meaning is the same. If English were written with logograms, there would be separate symbols for all of these morphemes, even the ones that share the same pronunciation or meaning.

Some of the oldest writing systems in the world, including Egyptian hieroglyphs and Sumerian cuneiform, made extensive use of logograms. Among living languages, Chinese and Japanese make the most extensive use of logograms. To a close approximation, we can say that each Chinese character represents a different morpheme. The three-morpheme word [tʃʊŋ]kwəʃɹənʃ [‘Chinese person’ is

represented by three characters (we use Simplified Chinese characters when discussing present-day Chinese):

(1) Chinese

我不是中国人。

我 不 是 中-国-人

wəʌ puʌ ʃiʌ tʃʊŋ]-kwəʌ]-ɹənʌ

I not be Middle-Country-person

'I am not Chinese.'

Chinese has several other morphemes pronounced [ʃiʌ] but with different meanings from the morpheme [ʃiʌ] in (1) (repeated in 2). For example, there is a homophonous noun that means 'thing'. As one would expect from logograms, this latter morpheme is written with a different symbol. Conversely, morphemes that are synonyms but with different pronunciations are also written differently from each other, as is the case with the two words for 'mouth' in (2).

- | | | | |
|-----|---|----------|---------|
| (2) | 是 | [ʃiʌ] | 'be' |
| | 事 | [ʃiʌ] | 'thing' |
| | 口 | [kʰəwʌ] | 'mouth' |
| | 嘴 | [tswəjʌ] | 'mouth' |

Unfortunately, logographies are never as pure as these examples would lead you to believe. It is by no means rare for the same symbol to represent two different morphemes. In Simplified Chinese, 谷 represents both the morpheme [kuʌ] 'grain' and the morpheme [kuʌ] 'valley'. In Japan, whose writing system is based on Chinese logograms, the symbol 食 can stand for [kuw] 'eat', [tabe] 'eat', [çoku] 'food', and [dʒoci] 'food'. It is easy to see the connections within these sets of morphemes. In the Chinese case, 谷 originally meant [kuʌ] 'valley' then was borrowed to represent another morpheme with the same pronunciation. In the Japanese case, the morphemes all have meaning in common, but not pronunciation. It would be tempting to theorize that Chinese script really represents sounds, not morphemes, because of cases like 谷, but the clear overall pattern really is that different morphemes are written with different symbols, even if pronounced the same. It would be equally tempting to hypothesize that Japanese is a semasiography, where symbols like 食 represent meaning rather than specific morphemes. But when a person reads aloud a Japanese text, the clear expectation is that she will read each of the Chinese logograms as a specific morpheme; it is by no means sufficient if she just paraphrases the meaning.

In both cases, it seems much more useful to take the point of view that the Chinese symbols are logograms, but with significant HOMOGRAPHY. Homography (Greek ὁμο- 'same') refers to a written form that is ambiguous. Such ambiguity is

arguably a pain for the reader, but in practice can almost always be overcome by taking context into account. It is found to some degree in most writing systems. How could so much homography arise in logograms? Sometimes there are specific historical circumstances. Much of the homography in Japanese resulted from the fact that it borrowed both its logograms and a huge amount of vocabulary from Chinese. It seemed natural to use the same Chinese logogram both for the Chinese words they borrowed as well as for Japanese words of the same meaning. There is also one major factor encouraging homography in all logographic writing systems: languages may have thousands of different morphemes. That huge number makes it difficult to invent and then to learn all the symbols that would be needed to uniquely represent each morpheme. When one begins to develop a logography, it must seem like an easy task. For the word *cat*, it is easy enough to draw a cat: 猫. The word *anger* is a little more abstract, but 😡 gets the point across. The great bulk of any language's vocabulary is not so easy to picture. Try picking a random paragraph from any book and inventing for each of its morphemes a logogram that you would feel confident passing on to your colleagues for their critical review. The simplest and most common solution to this problem is recycling – using an existing logogram for another morpheme that is in some sense similar to the morpheme that logogram already represents. In linguistic terms, that similarity almost always means similarity – though not necessarily identity – in sound or in meaning.

Of course, too much homography can be confusing. Many logographic writing systems address the problem with a touch of semasiography. A silent classifier is added to the logogram to help disambiguate the symbol by giving a hint as to the semantic domain of the intended morpheme. To take an example from Sumerian, the logogram 𒀭 was used to write [apin] 'plow' (noun), [engar] 'farmer', [uru] 'plow' (verb), and [absin] 'furrow'. When the word [apin] 'plow' was intended, the writer could optionally precede the logogram with a semasiogram for 'tree', signalling that he meant the 𒀭 that has something to do with trees, namely, an object made out of wood. When [engar] 'farmer' was meant, the writer could add the 'man' semasiogram. Chinese routinely did something very similar. To write the morpheme [ma] 'mother', Chinese borrowed the logogram 馬, which means [ma] 'horse'. Instead of tolerating the ambiguity, writers added the semasiogram 女 'female' to get 媽. Unlike Sumerian, Chinese writing actually fused the symbols together, so that it is visually obvious that 媽 is a single logogram representing a single morpheme. The great majority of Chinese logograms are formed in analogous fashion: borrowing a logogram for a morpheme that sounds similar to the morpheme for which a new symbol was needed, and fusing it with a semasiogram.

We mentioned earlier that there is a strong tendency for graphic symbols to be pictorial, but does that hold true for logograms? From the examples of Chinese logograms shown in this section, you might conclude not. But for Chinese and many other writing systems, very early forms of many logograms have been found that more clearly resemble their referent. The logogram for [xu] 'tiger' is now

written as in (3a), but one of its earlier attestations was (3b, cf. Sears 2016), a much more creditable picture of a tiger, especially if you tilt your head to the right. As time goes on, however, logograms change to become easier to write, more compact, and stylistically more consistent with each other. After a while, they may cease to be identifiable at all as pictures. Xiao & Treiman (2012) tested whether adults with no previous knowledge of Chinese could guess the meaning of 213 characters that evolved from simple pictures. Even though the task was very simple – choose which of two English words corresponds to the Chinese logogram – people’s guesses exceeded chance levels for only 15 of the logograms.

(3) Chinese ‘tiger’

a. 虎

b. 

2.1.1.3 PHONOGRAPHY

The other types of glottograms are sometimes referred to as PHONOGRAMS, because each symbol represents linguistic sound. Logographies and phonographies aren’t mutually exclusive; indeed, there are few if any logographies that don’t have a substantial phonographic subsystem. Sumerian and Japanese have large sets of phonograms that are used to spell out words. Most function words and grammatical affixes are spelled with phonograms, and, quite often, a morpheme is spelled with a logogram and phonograms at the same time. Technically this last usage is redundant, because a logogram already expresses sound, in that sound is part of a morpheme. But we have already noted how ambiguous logograms often are, so accompanying phonograms can be instrumental in clarifying which morpheme is meant.

2.1.1.3.1 Phonemography

The most efficient phonography, in terms of inventory size, is a PHONEMOGRAPHY: a set of symbols each of which represents a different phoneme. Hawaiian is an excellent example of such a script (4). Every phoneme of the spoken Hawaiian is represented by one phonogram in the Hawaiian orthography. Phonemographies whose characters represent consonants and vowels are called ALPHABETS.

(4) Hawaiian

ua mau ke ea o ka ‘āina i ka pono

ua mau ke ea o ka ?a:ina i ka pono

‘The country’s independence has rightly been preserved.’ (State motto of Hawai‘i, attributed to King Kamehameha III in 1843.)

Many phonemographies depart from the alphabetic ideal illustrated by Hawaiian. In English, stress is clearly unpredictable and therefore phonemic, but the writing system does not indicate it (5a). (We use angled quotes to refer to elements of spelling.) Sometimes one letter spells a sequence of phonemes (5b), or a phoneme is spelled by a sequence of letters (5c). Almost all phonemes have multiple spellings and almost all letters have multiple pronunciations, which most dramatically leads to the situation in which many homophones are spelled differently (5d).

- (5) a. ⟨pervert⟩: [pə¹vəɾt] (verb) or [ˈpəvəɾt] (noun)
 b. ⟨six⟩: ⟨x⟩ = [ks]
 c. ⟨think⟩: ⟨th⟩ = [θ]
 d. [saɪt]: ⟨sight⟩ ‘vision’, ⟨site⟩ ‘place’, ⟨cite⟩ ‘mention’

There are many reasons why phonemographies may not be perfectly one-to-one. One set of factors is historical: scripts behave in ways that may have made more sense in the past. English *sight* is spelled with a ⟨gh⟩ because it was pronounced with an extra consonant, [x], several centuries ago. Words borrowed from other literate languages usually preserve the spelling of the original: *site* is spelled like the stem of Latin *situs*, and *cite* is spelled like the stem of Latin *citare*. English [θ] has no letter of its own because its writing system was adapted from Latin, and Latin had no [θ] sound, so English made do by creating a DIGRAPH, a sequence of two letters that map to a single phoneme. This keeps English looking more like classical Latin than it would if a brand-new letter were used. Examples were here drawn from English because it has one of the more CONSERVATIVE (historically oriented) spelling systems, but such factors are by no means unknown in other writing systems, even those that undergo spelling reforms.

Another factor that is often repeated around the world is a reluctance to deal with SUPRASEGMENTAL phonemes or features. A phonetic feature is suprasegmental if it cannot be characterized as a separate segment of the speech stream, as a vowel or consonant can. In many languages in which stress is phonemic, such as English and Russian, it is not written. In Croatian, stressed vowels can be long or short, and pronounced with a rising or falling pitch; these distinctions are ignored in the standard writing.

One last class of problems we will mention here involves the fact that people tend to think in terms of meaning-bearing units of language – morphemes, words, and phrases – even when their basic task is to transcribe its phonemes. This has led to such nonphonetic conventions as leaving a space between words, even though there is usually no gap between them in speech. Especially in recent times, there is a certain tendency to always spell words the same way each time they are written. From there it is a small step to accepting that everybody in

the same speech community should always spell the same word the same way. Once people accept that this LEXICAL CONSTANCY is desirable, they may accept that words may have spellings they must memorize if they don't understand their motivation. When the rules of spelling vary depending on what word one is spelling, it begins to seem somewhat logographic. Nevertheless, it is clear enough that even in English, the basic symbols of the writing system represent phonemes, however imperfectly, and in Chinese, the basic symbols represent morphemes, however imperfectly.

2.1.3.2 Syllabography

Phonograms do not always represent individual phonemes. One major source of phonograms occurs when logograms get bleached of their semantic content and are used for their sound value alone. When that happens, the sound represented by the new phonogram could be as long as the original word or morpheme the logogram represents. In Ancient Egyptian, that could be at least as much as two syllables worth of sound. For example, the symbol  was originally the logogram for [xpr] 'beetle', then came to be used to represent the similar sequence [xpr] in other words, like [xpr] 'become'. The vowels of Egyptian are almost completely undocumented, which is why we haven't even guessed at them in our IPA transcription, but it is a safe bet that these words were two or three syllables long.

Chinese writing is the most thoroughly logographic of all writing systems, but even it makes use of phonograms in some contexts. Its approach to spelling foreign names and loanwords is to use logograms for their sound value alone. For example, one spelling of *Obama* is 奥巴马. When read off character-by-character, that is pronounced [aw\pa\ma\]. The pronunciation is much more apposite than the meaning of those characters, which would be something like 'mysterious hoped-for horse'. (The character 马 is the simplified version of the traditional character 馬 discussed earlier.) There is also an offshoot of Chinese script called *Nǚshū*, literally 'women's script', which was developed and used exclusively by women in Húnán. Although the script has not yet been thoroughly documented, it is clear that most of its characters were derived from Chinese logograms, but are used to represent only the sound of that logogram, not necessarily its meaning.

When Chinese or *Nǚshū* characters are used as phonograms, they represent an entire syllable, because Chinese morphemes are generally MONOSYLLABIC – one syllable long. Thus they are called SYLLABOGRAMS. A crucial part of the definition of a syllabogram is that it stands for a syllable holistically. A syllabogram can't be broken down into subsymbols, each of which represent the individual phonemes of the syllable. For example, in hiragana, one of two syllabographies used for Japanese along with Chinese logograms, the symbol for [ki] has nothing in common with those for syllables that start with [k] (6a) or end in [i] (6b).

(6) Japanese き [ki] versus:

a. [k-]

か く け こ

ka ku ke ko

b. [-i]

し ち に ひ み り

ʃi tʃi ni ʃi mi ɽi

If a writing system develops a separate syllabogram for each syllable in the language, the set of syllabograms can get quite large. A much more common structure of syllabographies is for them to contain symbols for only a small subset of a language's syllable types, usually the simplest and most frequent ones. Syllabograms for CV sequences – a consonant followed by a vowel – are very common. Sumerian was unusual in also having VC syllabograms. Still, the spoken Sumerian language had CVC syllables, but the writing system had no CVC syllabograms.

Such mismatches are dealt with in a variety of ways by different systems. Sometimes syllabograms are accompanied by non-syllabographic characters that extend their expressive power to accommodate more complex syllables. A good example of this approach is afforded by Japanese. Its syllabograms do not indicate voicing, vowel length, CODAS (consonants that follow the vowel in the syllable), or palatal consonants, but in modern usage, these ambiguities are almost entirely resolved by a variety of devices, such as using diacritics – small auxiliary symbols used to distinguish ambiguous referents. Japanese also has a single phonemogram, ん [n], which expresses its most common syllable coda. Without such extensions, Japanese would need a huge number of syllabograms to represent all its syllable types; instead, each of its two syllabaries comprises only 44 distinct syllabograms.

(7) Japanese extended syllables

き ぎ きい きん きゃ

ci ci-VOICED ci-i ci-n ci-ja

[ci] [ji] [ci:] [cin] [ca]

The other major way of using a small number of syllabograms to represent a large number of syllables is UNDERREPRESENTATION. That means that a syllabogram or sequence of syllabograms can stand for two or more values, usually values that are similar to each other. For example, in Sumerian, a CVC syllable such as [nan] could be spelled by writing two phonograms in a row: [na] + [an]. Many individual syllabograms were ambiguous. Often they didn't specify the height of the vowel or the voicing of the consonant. For example, the same symbol stood for either [si] or [se], and the symbol for [zu] also represented [su]. In Cherokee, a typical syllabogram can be pronounced 24 different ways (Montgomery-Anderson

term *ABJAD* as a more convenient if less morphologically transparent term for a consonantal phonemography. But Gelb does have a point in that if one reads off a Phoenician text character by character, most of the characters will be sounded as a consonant-vowel sequence, just as when one is reading syllabograms. Therefore we compromise with Gelb and characterize this system as a type of syllabically organized phonemography.

2.1.3.4 Abugida

In perhaps all syllabically organized phonemographies, consonants are treated as more basic entities than vowels are. An *abjad*, where vowels are totally unrepresented, is an extreme case. In other such scripts, vowels are represented, but are graphically subordinated to any preceding consonant. Daniels (1990) has bestowed the name *ABUGIDA* on such systems. A typical abugida is the Gujarati script of India, illustrated in (9). Below the sentence presented normally, we have written each character separately, with its pronunciation below it in IPA.

(9) Gujarati

તાડપત્ર પર લખવાની બે પદ્ધતિઓ હતી

તા ડ પ ત્ર પ ર લ ખ વા ની બે પ દ્ધ તિ ઓ હ તી
 ta ɖ pə trə pə r lə kʰ va ni be pə ddʱə ti o hə ti

‘There were two systems of writing on a palm leaf.’ (Mistry 1996)

The main difference between this abugida and an *abjad* is that diacritics are added to the basic consonant phonemograms to indicate what vowel follows the consonant. બે [be], for instance, is the consonant બ [b] with a mark above it to show that the consonant is followed by the vowel [e]. Different diacritics are placed on different sides of the base consonant: બુ [bu] illustrates a diacritic written below the base, બી [bi] contains a vowel that extends to the right, and બિ [bi] contains a vowel that extends to the left. (The last two are homophonous because the writing system preserves an archaic distinction of vowel length.) Even the symbol for [a], as in બા, is a diacritic attached to the બ, despite the fact that its spacing and position make it look like it could be a character itself, albeit an unusually thin one.

Most abugidas treat one vowel as a default or *IMPLICIT VOWEL*. That vowel is not represented by a diacritic. In Gujarati, the implicit vowel is [ə]. પ is the basic phonemogram for [p], but in (9) it spells [pə] several times without any added diacritic. In principle, any consonant symbol is supposed to be read with the implicit vowel unless there is some graphic clue to the contrary, such as another vowel’s diacritic. All of the [ə] sounds in our text are implicit vowels, not represented by any explicit symbol. But what about consonant clusters? If in trying to spell [ddʱə], for example, we had written દ [d] followed by ઢ [dʱ], the sequence could spell [dədʱə] because of the implicit vowel rule. To spell a consonant sequence, many abugidas

use consonant LIGATURES, or CONJUNCTS. The two consonants are combined so that they look like a single complex consonant. In this case $\epsilon + \text{ʌ}$ form ʌ . This conjoining is the graphic clue that no vowel is pronounced after any consonant except the last one in the conjunct. Most abugidas also employ, alongside conjunction, the simpler solution of having a diacritic that means ‘no vowel’, and some writing systems, such as Tamil, use that approach exclusively.

If you have been studying (9) closely, you might at this point object that you see several instances in which adjacent consonant sounds are not represented by a conjunct. [taɖpətrə] ‘palm leaf’, for example, isn’t written with conjoined ʌ and ʌ symbols. The reason for this omission is that there used to be an [ə] there. The script is conservatively spelling the older form [taɖəpətrə]. Many modern Indian languages have undergone SYNCOPE (loss from the interior of a word) as well as APOCOPE (loss from the end of a word) of [ə] or the historically corresponding vowel, with the consequence that the rules about implicit vowels and consonant conjunction can be fairly difficult to apply to the current language.

Because most of the characters in an abugida contain one vowel, whether diacritic or implicit, they can mostly be pronounced as syllables. While it is true that abugidas are syllabically organized, not too much should be inferred from that fact. For one thing, not every character corresponds to a syllable of speech. Even if we restore all the deleted [ə] sounds, there are some egregious mismatches between written character and spoken syllable. For example, [pədd^hətio] ‘systems’ is spelled with the character ʌ (9), but [dd^hə] would be a pretty horrible syllable; the true syllabification is [pəd.d^hə.ti.o]. True, most characters do end up spelling syllables, but the overarching principle is to group each vowel with however many consonants immediately precede it (schematically, C*V), regardless of whether that creates a linguistic syllable. It should also be noted that the characters of an abugida are not syllabograms, because they are composed of individual symbols, each of which denote individual phonemes in the syllable. In a true syllabography like hiragana (6), [be], [bu], [bi], and [ba] would look no more similar to each other than they would to any other syllabogram. Thus abugidas are phonemographies.

2.2 Genetic classification of writing systems

A genetic classification of writing systems groups together systems that have a common origin. Historians interested in the cladistics of writing systems do not have a standard technique that can be summarized as succinctly as the comparative method used for the cladistics of languages. There are, however, a few general principles that are helpful to keep in mind as we look at languages and their writing systems in the following chapters.

2.2.1 GENERAL PRINCIPLES

2.2.1.1 The spread of scripts is sociological

The spread of scripts has very little to do with linguistic properties of the languages they represent. The cladistics of a language's writing system does not always match that of the language itself, and both do not necessarily match the genetics of the people who speak and write them. The fact that people use the same script for writing two different languages does not mean that the people or the languages are genetically related to each other. For example, the Latin script you are reading here is used for many related (Indo-European) languages and for thousands of unrelated languages throughout the world, from Zulu in Africa to Quechua in South America to Vietnamese in Asia. Keeping this principle firmly in mind, however, borrowing does take place most freely between groups with cultural affinities of one sort or another. To take one example, the Arabic script, because of its use in writing the Koran, is strongly associated with Islam, and therefore has been extensively used for writing languages spoken predominantly by Muslims, including the Afro-Asiatic language Arabic, the Indo-European language Persian, and many others. On the other hand, many other predominantly Muslim countries shifted away from the Arabic script to the Latin or Cyrillic alphabet during the course of the twentieth century. In general, changes in writing systems result from political, economic, and other social factors. Purely linguistic factors, such as genetic or typological relations among languages, are rarely, if ever, decisive.

2.2.1.2 Scripts change

Like language and every other human institution, scripts change. There are some well-known cases in which the writing medium itself clearly influenced the change. The Sumerian cuneiform script started out pictorial but ended up taking the form of a collection of wedge-shaped marks. This change must be due in large part to the Sumerians' choice of writing medium: it is difficult to drag a writing instrument through clay to draw a picture, but easy to poke clay with a wedge-shaped stylus. But the role of writing medium as an agent of change can be overstated. If there is a single factor that most often drives change in a particular direction, it is *ECONOMY*. Writing systems that start off with recognizable pictorial symbols have a strong tendency to simplify them over time, making them less recognizable, but much easier to write. This same factor of economy also tends to make the different symbols of a writing system increasingly resemble each other over time: it appears to be easier to write large amounts of text if similar motor strategies are used for the different symbols. Because of this strong and constant tendency toward changing the graphic form of writing, it is always useful to keep an open mind when considering whether formally different scripts may be related.

2.2.1.3 Scripts are similar

The principle of economy also often causes some characters in one script to look like characters in another script. After all, there are only so many shapes that are both easy to write and also contrast adequately with the other symbols in the same script. This natural tendency toward similarity can make it difficult to decide whether scripts are related, in the absence of historical trails connecting them. Because there is no comparative method to help decide whether shapes show up in two different scripts by mere chance, it is hard to know how to evaluate claims that graphic similarities point to common genetic origin. To take one example, some Canaanite inscriptions have one symbol that is a simple square shape. Some authors have pointed to a square-shaped symbol in some Egyptian inscriptions as evidence that Canaanite is based on Egyptian hieroglyphs (e.g. Goldwasser 2010). The square, of course, is only part of the evidence that these authors adduce, but questions arise. How often might we expect inventors of totally unrelated scripts to independently come up with the idea of using a square? And if a single square is weak evidence, just how many similar pieces of weak evidence are required to add up to a substantive case? In general, shape alone is a tenuous basis for drawing connections between scripts; the case becomes strong only when symbols with the same shape have the same function in the two scripts.

2.2.1.4 Glottographies often evolve from semasiographies

Readers often puzzle over wild discrepancies in claims about the age of a writing system, or the number of writing systems discovered in an area. Much of the reason for such disagreements is a spotty historical record. But one cause for disagreement that often goes unspoken boils down to a matter of definition: What is writing? Many scholars say that any system of using conventional symbols to communicate specific ideas is writing (Boone 2004); this definition would include semasiographies. Others, especially linguists, equate writing with glottography. By this stricter definition, writing systems in the ancient Near East, including the Sumerian writing tradition, arguably do not manifest themselves as such until there is clear evidence that texts are meant to be read off as a specific string of words, which appears to have happened around the 28th century BC (Cooper 2004). Prior to that time, Sumerian had a well-developed semasiography, which was dedicated largely to keeping precise records about matters such as the size of grain inventories. Semasiographic Sumerian writing may go back five more centuries. The archaeologist Schmandt-Besserat (2002) even believes that a few Sumerian logograms can be traced to clay counting tokens used as early as 7500 BC. Clearly the correct dates for the beginning of Sumerian writing depend heavily on the data available to researchers and on its correct interpretation, but readers also need to understand precisely what individual authors mean by *writing*.

2.2.2 POLYGENESIS OF WRITING

Nobody knows how many times the idea of writing independently arose and bore fruit. Table 2.2 is one liberal attempt to address this question. Scripts are indented under the scripts that they were known to evolve from. Scripts that aren't indented are essentially independent creations. Thus, knowing Ogham will do you absolutely no good in reading Linear B.

Nevertheless, a skeptic might wonder whether the unindented scripts are all truly completely independent of each other. Many of these scripts have been

TABLE 2.2

Scripts of the world

Script	Type	Place
Ogham	alphabet	Ireland
Cretan		
· Linear B	syllabography	Greece
Glagolitic	alphabet	Croatia
Armenian	alphabet	Armenia
Georgian	alphabet	Georgia
Luvian hieroglyphs	logography	Turkey
Cuneiform	logography	Iraq
Ugaritic	abjad	Syria
Old Persian	alphabet	Iran
Canaanite	abjad	Lebanon (see Table 2.3)
Orkhon	alphabet	Mongolia
Oj Chiki	alphabet	India
Pahawh Hmong	alphabet	Laos
Chinese	logography	China
· Nǚshū	syllabography	China
· Hiragana	syllabography	Japan
· Katakana	syllabography	Japan
Yi	syllabography	China
Hangeul	abugida	Korea
Hieroglyphic	logography	Egypt
· Hieratic	logography	Egypt
· Demotic	logography	Egypt
Berber	abjad	Tunisia
· Tifinagh	alphabet	Morocco
Vai	syllabography	Liberia
N'ko	alphabet	Guinea
Cree	abugida	Canada
Cherokee	syllabography	United States
SignWriting	visual iconography	United States
Maya	logography	Guatemala

developed in modern times, so that we can presume their inventors had at least seen writing. In fact, in some cases we don't have to presume: we have certain biographical knowledge. N'ko, for example, was developed by Souleymane Kanté, a teacher and bookseller who is known to have read widely in several languages. He invented the script as an explicit indigenous challenge to scripts brought to Africa by the colonial powers. Because writing is a fairly sophisticated idea – communicating by visually mapping the symbols of an aural communication system – one might wonder whether the development of new writing systems depends on what the anthropologist Kroeber (1940) called *STIMULUS DIFFUSION*. Perhaps one Early Bronze Age genius in Sumeria or Egypt came up with the idea of glottography, and the inventors of all other scripts were inspired by an already existing script whose lineage traces back to that original writing system.

Is it just a coincidence that Egyptian hieroglyphs and Sumerian cuneiform developed during the same era (by 2700 BC) and not terribly distant from each other? Suspicion grows when one considers how similarly the two scripts worked. They both started out as logographies, with most of the logograms originating as realistic pictures of the object named by the morpheme. Both based a large number of their logograms on other logograms with similar sounds, controverting the morphemic basis of logograms and elevating the phonographic principle. Both of them invented ways to spell out words on a purely phonographic basis, which they could use instead of logograms or in conjunction with logograms, often redundantly spelling out some or all of the sounds of the morphemes. Both of them had a restricted set of symbols that could be used as silent semasiograms that semantically classify words. Logograms, phonograms, and semasiograms did not differ formally from each other in the two scripts, and indeed the same symbol could in different contexts be used in two or more of these ways.

Yet even all these similarities do not necessarily mean that Egyptian hieroglyphs and Sumerian cuneiform have a common origin. The Maya script of Mesoamerica has all of these same properties (Mora-Marín 2008). Given the total unlikelihood of a causal connection between the Ancient World scripts and the Maya script, we conclude that it is quite possible that Egyptian hieroglyphs and Sumerian cuneiform also developed independently of each other. The most reasonable conclusion from these three sets of developments is that basing a writing system on morphemes is a natural idea, but one that is difficult to elaborate fully without adding some phonography. The case of Maya makes less credible the idea that the invention of new scripts depends on stimulus diffusion.

There is much more we can learn from studying the hundreds of other scripts and writing systems of the world. Table 2.2 is our list of what we consider the most important families of scripts, and we encourage you to explore them further, starting with books such as Daniels & Bright (1996) and online sources such as Wikipedia and Omniglot (Ager 2015).

Keep in mind that the same script can be used for many different languages. The application of a script to a specific language constitutes a *WRITING SYSTEM*.

The ways scripts are used and adapted for different languages can be as varied as the differences between scripts themselves. The application of the Chinese script to various languages is particularly complex. It is still used for writing Chinese logographically, in much the same way that it was used in the Shāng dynasty over 3,000 years ago. The main development over the years has been in the shapes of the individual characters. The latest development in that evolution occurred in the mid-twentieth century, when the government of the People's Republic of China sponsored a massive program to simplify the official shapes of many of the characters. Chinese script was also borrowed by many other cultures within the Chinese sphere of influence, being adapted for such languages as Japanese, Korean, and Vietnamese. At present, Chinese logograms are still fundamental to Japanese, less so to Korean, and are rarely used in Vietnamese. In Japan, two different syllabographies were developed by using certain logograms solely for their phonetic value, then simplifying the forms of those characters. In Korea, an abugida was invented in the fifteenth century. Many people in Japan and Korea, however, feel that their languages have so much homophony – due primarily to heavy borrowing of simple one-syllable morphemes from Chinese – that logograms still play an essential role in resolving ambiguity.

The vast majority of languages today use a script that descended from the Canaanite script (Table 2.3). Many scholars emphasize the possible influence of the Egyptian writing system on Canaanite, pointing out that some Canaanite symbols look like Egyptian hieroglyphs. The counterargument is that the Canaanite symbols are very simple pictures or geometric shapes, which could easily have developed independently of any existing model. Furthermore, the similar symbols do not have the same function in the two scripts. A better argument is that Canaanite, like Egyptian, did not write vowels. Their shared lack of symbols for vowels is striking – there are very few cases of scripts independently inventing the idea of ignoring vowels – but in the case of Canaanite, it has a simple explanation. The Canaanite characters are ACROPHONIC: each of them is a picture whose phonetic value is the first sound in the word for the pictured object. Thus a picture of an ox, which was called approximately [ʔalp], was used to represent the sound [ʔ]. Because in Canaanite languages no word could begin with a vowel, no vowel letters could be formed by this process. The lack was unfortunate, but most writing systems are defective in some respect; people can deal with a surprising amount of ambiguity when reading texts written in their own language.

The Canaanite script was the ultimate source of a large number of scripts, most of which, in the first instance at least, inherited its lack of vowel letters. The Ge'ez alphabet, used to write Amharic and other languages of Ethiopia, eventually developed diacritics to indicate which vowel should follow the consonant. For example, the word Ge'ez [gɪʕiz] is itself spelled (from left to right) ጊሕዝ, where the first two letters are modified from their basic forms ጊ [g] and ሕ [ʕ], respectively, in such a way as to indicate that the consonants they represent are each followed by the vowel [i]. Many other descendant scripts, such as Phoenician,

TABLE 2.3

Canaanite family of scripts

Script	Type	Place
Canaanite	abjad	Lebanon
· Phoenician	abjad	Lebanon
· · Greek	alphabet	Greece
· · · Etruscan	alphabet	Italy
· · · · Latin	alphabet	Italy
· · · · · Runic	alphabet	Denmark
· · · · Cyrillic	alphabet	Russia
· · · · Coptic	alphabet	Egypt
· · · Gothic	alphabet	Bulgaria
· · · Aramaic	abjad	Iraq
· · · Syriac	abjad	Iraq
· · · · Sogdian	abjad	Tajikistan
· · · · · Old Uyghur	alphabet	Mongolia
· · · · · · Mongolian	alphabet	China
· · · · · · Manchu	alphabet	China
· · · Hebrew	abjad	Israel
· · · Arabic	abjad	Egypt
· · · · Dhivehi	abugida	Maldives
· · · Mandaic	alphabet	Iran
· · · Brahmi	abugida	India
· · · · Devanagari	abugida	India
· · · · Gujarati	abugida	India
· · · · Bengali	abugida	Bangladesh
· · · · Tibetan	abugida	China
· · · · Gurmukhi	abugida	India
· · · · Oriya	abugida	India
· · · · Sinhala	abugida	Sri Lanka
· · · · Kannada	abugida	India
· · · · Telugu	abugida	India
· · · · Tamil	abugida	India
· · · · Malayalam	abugida	India
· · · · Pallava	abugida	India
· · · · · Mon	abugida	Myanmar
· · · · · Javanese	abugida	Indonesia
· · · · · Baybayin	abugida	Philippines
· · · · · Khmer	abugida	Cambodia
· · · · · · Thai	abugida	Thailand
· · · · · · Lao	abugida	Laos
· · · · · · Cham	abugida	Cambodia
· · · Pahlavi	abjad	Iran
· · · · Avestan	alphabet	Iran
· · Samaritan	abjad	Israel
· South Arabian	abjad	Yemen
· · Ethiopic	abugida	Ethiopia

Aramaic, and the latter's descendants Arabic and Hebrew, make some of their consonants do double duty as vowels (e.g. using [w] to stand for [u]), but still get by without writing many if not most of the vowels in a text. Yet when the Greeks borrowed the Phoenician script, they somehow ended up with a full complement of vowel letters. This is traditionally celebrated as a stroke of sheer genius (Gelb 1963), but it may also be the straightforward result of reinterpreting the Canaanite/Phoenician letter names in a new phonology. The Greeks may not have recognized the Phoenician consonant sounds absent from the Greek language and therefore took the second sound in the name of the letter to be the letter's value. Thus, for example, the Phoenician letter that was called [ʔalpu] 𐤀 represented a glottal stop, a sound that did not exist as a phoneme in Greek. It was very natural, then, for a Greek learner to assume that [ʔalpu], whose name was adapted as [álp^ha] Α in Greek, represented the sound of the vowel [a]. Similarly, Phoenician letters representing various pharyngeal (articulated in the throat, like [ʕ] and [ħ]) and glottal sounds ([h]) not found in the inventory of Greek phonemes ended up being interpreted by the Greeks as representing other vowel sounds. However, it may be that our scenario underestimates the Greeks: they may have consciously decided to repurpose the “useless” Phoenician letters for the vowel sounds.

Eventually, the Greek alphabet, with its vowels, spread to the Romans, who spread it all over Europe as the Latin alphabet. In turn, European colonizers and traders spread it over much of the rest of the world. In addition, the Greek alphabet was the basis of the Cyrillic alphabet, used primarily by the Slavic peoples who were Christianized by the missionaries from Byzantium. Cyrillic is also used by many other languages spoken in or near predominantly Slavic-speaking countries, or in the former Soviet Union.

Some descendant of Canaanite, probably a form of Aramaic, seems to have provided the basis for the Brahmi script in India, which, in turn, is the basis of most of the current scripts of South Asia: several Indian scripts, as well as those used for Burmese, Cambodian, Thai, Laotian, and also Tibetan. All of these scripts are abugidas.

2.3 Exercises

2.3.1 LOGOGRAPHY AND PICTOGRAPHY

Devise a purely logographic script to write the following English passage:

The powerful chief of the Blackfeet attacked the invaders at the mountain pass and defeated them. He took 200 prisoners, including Chief Eagle Feather.

Next, try to convey the gist of the preceding message using pictorial representation. Instead of actually drawing the picture, you may just describe what kind of picture or series of pictures you might draw. How might you convey the proper names? If you run into any difficulties in accomplishing the assigned task, be sure to explain in detail what they are.

2.3.2 TYPES OF WRITING SYSTEMS

Examine the three different scripts given below and determine what type of writing is represented by each. Briefly explain the reasoning that led you to your conclusion in each case. Both the writing systems and the languages are made up; if you recognize the symbols, ignore whatever you know about them in real life. The first line is script, and the second line is a phonemic transcription. Each word has one morpheme.

- (1) ነጋጽ ሠጋሩ ሰጠኝ ነጋጽ ጸ
 kamani wikara umikama ni
 'Did Kamani bring his sword?'

- (2) ተ ለ ገጠ ለ ል ል ጸ
 kad taka ugura taka maridu marakat
 'When did the man take the donkey?'

- (3) ጠጠጠጠጠጠ ጠጠጠጠጠጠ ጠጠጠጠጠጠ
 mangetukan warganut kereto
 'Virtue is always triumphant.'

2.3.3 AIGA DOT SEMASIOGRAMS

Figure 2.3 contains some additional examples of semasiograms designed by the AIGA to convey messages to travelers in transportation hubs such as airports. See if you can decipher the message of each semasiogram. Give a brief statement about your reasoning in each case. Do you think that all of the semasiograms would be understood by all people? For example, would an ordinary villager from a third world country who has only a fourth grade education understand them? Why, or why not?

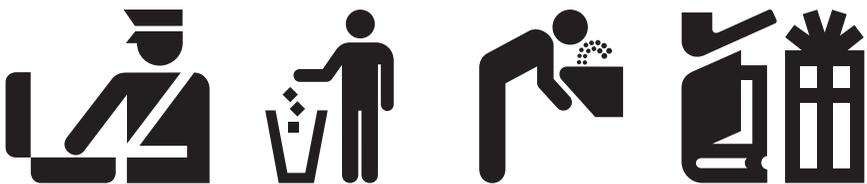


FIGURE 2.3 AIGA DOT semasiograms, for Exercise 3. Courtesy of AIGA.

2.4 Suggested readings

- ✘ *The Blackwell encyclopedia of writing systems* (Coulmas 1996). An authoritative description and illustration of more than 400 writing systems in encyclopedia format.
- ✘ *The Chinese language: Fact and fantasy* (DeFrancis 1984). The main purpose of this book is to debunk some of the myths that have arisen in connection with Chinese writing.
- ✘ *The first writing: Script invention as history and process* (Houston 2004).
- ✘ Fundamentals of grammatology (Daniels 1990). This short article is a good starting point for studying writing systems, because it presents a helpful typological schema and some influential views on the historical development of writing.
- ✘ Signs of life (Schmandt-Besserat 2002). A succinct summary of the author's theory of the prehistory of cuneiform.
- ✘ *The story of writing: Alphabets, hieroglyphs & pictograms* (Robinson 2007). Introductory-level overview of major types of writing, with many illustrations.
- ✘ *The world's writing systems* (Daniels & Bright 1996). This is the most up-to-date comprehensive reference on the subject.
- ✘ *Visible speech: The diverse oneness of writing systems* (DeFrancis 1989). This work challenges many traditional concepts concerning the nature of writing, especially logographic writing.
- ✘ *Writing systems* (Sampson 2015).
- ✘ *Writing systems: A linguistic approach* (Rogers 2004). An excellent textbook covering the history of the world's major writing systems, their structure, and the social context in which they evolved.

Europe

Table 3.1 is a summary list of the languages of Europe. Because of the large numbers of languages in Europe, the first column generally lists a language clade, and the second column directs you to the part of this book where the individual languages are discussed. Occasionally we find it convenient to discuss together multiple language families that are spoken in the same region; such groupings are identified in our tables as an *area*, to forestall the impression that they are known to be a linguistic clade. The individual languages Basque and Turkish are listed separately because they do not form any natural groupings with other European languages. Basque is not related to any other language, and Turkish is related to the Turkic languages of Asia and, thus, is discussed in Chapter 4.

The second column of the table tells which section of this book discusses the language or group. The “Size” column is a rough estimate of the number of people who speak the languages fluently and use them on a regular basis. To save space and to make it easier for you to take in orders of magnitude at a glance, we give the number of digits that it would take to represent the number of speakers. For example, “7” in the row for the Uralic family means that the number of speakers is between 1,000,000 and 9,999,999. It is remarkably difficult to get accurate figures for the number of speakers, and available estimates are often not commensurable between languages, so these numbers should be taken with a grain of salt. The last column tells where the languages are spoken. Of course, a large number of languages are spoken in diverse locations; in many cases, the location given is selected from many candidates as the one most closely associated with the language.

The three-letter codes next to the language names are ISO 639-3 codes (SIL 2015), the product of an international effort to establish unique identifiers for all human languages. The codes are more succinct and less ambiguous than language names and will help you look up additional information about languages. For example, to find more information about Basque in *Glottolog*, you can direct your browser to the URL <http://glottolog.org/glottolog?iso=eus>; for other languages, substitute their ISO code for the final three letters of that URL. In the Symbols

TABLE 3.1
Languages of Europe

Name	Section	Size	Location
Indo-European family	§3.1	10	Eurasia
Uralic family	§3.2	7	N Eurasia
Caucasus area	§3.3	8	SE Europe
Basque (EUS)	§3.4	6	SW Europe
Turkish	§4.1.1	8	Turkey

TABLE 3.2
Some Indo-European languages

Name	Section	Size	Location
Celtic	§3.1.1	7	Britain
Germanic	§3.1.2	9	NW Europe
Italic	§3.1.3	9	S Europe
Greek (ELL)	§3.1.4	8	Greece
Albanian (SQI)	§3.1.5	7	Albania
Balto-Slavic	§3.1.6	9	E Europe
Armenian (HYE)	§3.1.7	7	Armenia
Indo-Iranian	§3.1.8	10	SW Asia
Anatolian	§3.1.9	0	Turkey
· Hittite (HIT)		0	Turkey
· Luwian (HLU)		0	Turkey
· Lycian (XLC)		0	Turkey
Tocharian	§3.1.10	0	China
· Tocharian A (XTO)		0	China
· Tocharian B (TXB)		0	China

and Abbreviations section we list in alphabetical order all the ISO codes that we use in this book, along with the corresponding language name. The codes are also included in the summary tables and are used on our maps.

3.1 Indo-European

The vast majority of the languages currently spoken in Europe belong to the Indo-European family, which is one of the most extensively studied language families in the world. Here we discuss in some detail the branches of the Indo-European language family, in the order summarized in Table 3.2. Figure 3.1 uses ISO codes to show the location of several of the Indo-European languages before their post-Columbian expansion from Europe and Asia into Africa, the Americas, and Australia.

3.1.1.1 CELTIC

Table 3.3 lists the main living Celtic ([^lseltɪk] or [^lkeltɪk]) languages. In ancient times, the Celtic branch had a very wide distribution. Inscriptions in Celtic languages, some dating back to 600 BC, have been found as far east as Turkey. At present, however, Europe has only small pockets of Celtic speakers in the north-westernmost parts of the continent because of pressure from competing languages.

The existing Celtic languages comprise two groups: Brythonic and Goidelic. The Brythons were the Celts who settled in Great Britain in the Iron Age. With the arrival of the Germanic tribes who brought the Anglo-Saxon culture to Great Britain, many Brythons relocated to northwest France, where the descendant of their language is known as *Breton*. In Great Britain itself, the remaining Brythons were increasingly marginalized to the west of the island, where today their language survives primarily in the form of Welsh, although a campaign to revive Cornish has had some success.

The Goidels were Celts who settled in Ireland at least three millennia ago. They subsequently carried their language to Scotland and the Isle of Man, which led to three national variants, usually referred to as Irish, Scottish Gaelic, and Manx. In English, modern Goidelic languages are sometimes referred to as Gaelic, which is similar to the ENDONYMS, the names for the languages in the languages themselves. The Irish word for Irish is *Gaeilge*, the Manx word for Manx is *Gaelg*, and the Scottish Gaelic word for Scottish Gaelic is *Gàidhlig*.

The earliest inscriptional evidence of Celtic in the British Isles is in the form of OGHAM inscriptions, which recorded an early form of Irish. This unusual script consists entirely of straight lines that were carved on the edge of a stone. Letters were distinguished from each other by factors such as the side of the edge from which they branched off, and the number of lines carved. Using a horizontal line to represent one edge of the stone, which was normally vertical, π ππ ππ πππ spells out letters for the sounds [l], [w], [s], and [n], in that order, whereas strokes carved on the opposite side of the stone π ππ ππ πππ spell [d], [t], [k], and [kʷ] (McManus

TABLE 3.3
Some Celtic languages

Name	Size	Location
Brythonic	6	
· Breton (BRE)	6	France
· Cornish (COR)	4	England
· Welsh (CYM)	6	Wales
Goidelic	6	
· Irish (GLE)	6	Ireland
· Scottish Gaelic (GLA)	5	Scotland
· Manx (GLV)	2	Isle of Man
Gaulish (XTG)	0	France

1991: 2). There were also sets of letters whose strokes crossed the edge line, either perpendicularly or at an angle. Unfortunately, the ogham inscriptions were all very short, consisting almost entirely of personal names in the genitive case: the name of the person who, in some sense, owned the stone.

The modern Celtic languages are mostly VSO, although this order is often obscured by PERIPHRASTIC verbal constructions, which use auxiliary verbs in addition to lexical verbs. In Irish, the auxiliary verb comes first, but the main content verb follows the subject (1):

(1) Irish

Tá sí ag crú na mbó.

tɑ: ʃi: ə kru: nə mo:
be(AUX)[ipfv] PRO(3SG.F)[nom] at milk[ger] DEF.ART[GEN.PL] GEN.PL\COW

‘She is milking the cows.’ (Christian Brothers 1993: 129)

Parentheses in glosses enclose grammatical categories that are inherent in the lexeme.

AUX = AUXILIARY VERB, a verb that primarily imparts grammatical information.

IPFV = IMPERFECTIVE ASPECT, denoting an action in progress (see further §3.5.4.5).

F = FEMININE GENDER, a class of words that includes those referring to women.

PRO = PRONOUN.

GER = GERUND, a verbal form that behaves in some respects like a noun.

GEN = GENITIVE CASE, which notes a grammatical dependency on another noun (here it is the object of a gerund). Backslashes in glosses indicate affixes that are not linearly separable from their base.

Sentence (1) could be rendered very literally as ‘Is she at milking of the cows.’ The main verb has been nominalized and turned into the object of a preposition, and its own object is expressed in the genitive case.

The last word in (1) illustrates APOPHONY, a process by which a grammatical category is expressed by a partial change of pronunciation rather than by adding an affix. The word for ‘cows’ in the genitive is normally *bó* [bo:], but any word that begins with [b] and follows the definite article *na* in the genitive plural must change the [b] to [m]; note how the spelling ⟨mbó⟩ expresses both the surface pronunciation and the original sound. Apophony affecting consonants is also referred to as a MUTATION. Celtic languages typically have two or three different series of mutations, which have different outcomes for different consonants. The change of [b] to [m] is an example of NASAL MUTATION.

3.1.2 GERMANIC

The Germanic languages are spoken in northwest Europe, from Austria to Iceland (Table 3.4).

TABLE 3.4
Some Germanic languages

Name	Size	Location
Northwest Germanic		
· North Germanic	8	NE Europe
· · West Scandinavian		
· · · Old Norse (NON)	7	Norway
· · · Norwegian (NOR)	7	Norway
· · · Faroese (FAO)	5	Faroe Islands
· · · Icelandic (ISL)	6	Iceland
· · East Scandinavian		
· · · Danish (DAN)	7	Denmark
· · · Swedish (SWE)	7	Sweden
· West Germanic	9	W Europe
· · Anglo-Frisian		
· · · English (ENG)	9	British Isles
· · · Western Frisian (FRY)	6	Netherlands
· · Low Germanic		
· · · Dutch (NLD)	8	Netherlands
· · · Afrikaans (AFR)	7	South Africa
· · High Germanic		
· · · German (DEU)	8	Germany
· · · Yiddish (YDD)	7	Poland
Gothic (GOT)	0	Bulgaria

Until 1818, the Germanic languages were disconcerting to historical linguists, because their vocabulary did not seem to correspond very closely to that of other Indo-European languages. For example, the word for ‘foot’ was [pəd-] in Sanskrit (the hyphen indicates we are omitting the inflectional suffixes), [ped-] in Latin, and [pod-] in Greek, but in the Germanic languages it is English [fʊt], German [fu:s], Icelandic [fʊt^h-], and so forth. Virtually all OBSTRUENT CONSONANTS in the Germanic languages – sounds in which the airflow is significantly blocked, namely, stops, affricates, and fricatives – are different from those seen in cognates in any other Indo-European language. But Rasmus Rask worked out that, even though the obstruents do not match between the two sets of languages, they do correspond. That is, [p] in Sanskrit, Latin, and Greek words corresponds to Germanic [f] not only in *foot* but regularly in other words as well, such as ‘father’: Sanskrit [pitər-], Latin [pater], Greek [paté:r], compared to English [faðr], German [fa:tɐ], Icelandic [faðr], and so forth. Further, a simple set of three rules known as GRIMM’S LAW accounts for the basic correspondences with all the Indo-European obstruents:

(2) Grimm's law

- a. voiceless stops became fricatives, e.g. [t] > [θ]
- b. voiced stops became voiceless, e.g. [d] > [t]
- c. breathy-voiced stops acquired normal voice, e.g. [d̪] > [d]

BREATHY VOICE describes a type of sound made with the vocal folds vibrating but letting much more air through than for normal voicing. Note that the three types of changes in Grimm's law occurred in the order listed. Each successive change replaced the consonants that were lost by the previous change, so that Germanic did end up with voiced and voiceless stops, though not breathy-voiced stops.

The Germanic languages are SVO in main clauses, but with a twist. Most of them are actually "main verb second". In (3a), the subject comes before the verb and the object comes immediately after the verb. But (3b) shows that if any other element is placed at the beginning of the sentence, the subject must be placed after the verb: the main verb must always come in second position. It is even possible in many Germanic languages to use the verb-second order OVS: *Den Brief schreibt der Schüler*.

(3) German

- a. Der Schüler schreibt einen Brief.

de:-ə	ʃy:lə	ʃʁaɪp-t	aɪn-ən	bʁi:f
DEF.ART-NOM	student	write-3SG	INDEF.ART-ACC	letter
	S	V		O

'The student writes a letter.'

INDF = INDEFINITE REFERENCE.

- b. Heute schreibt der Schüler einen Brief.

hɔɪtə	ʃʁaɪp-t	de:-ə	ʃy:lə	aɪn-ən	bʁi:f
today	writes-3SG	DEF.ART-NOM	student	INDEF.ART-ACC	letter
	V		S		O

'Today the student writes a letter.'

A notable typological feature characteristic of most Germanic languages is a very large number of vowels. Germanic languages often have sets of vowel phonemes that have very small differences in place of articulation, particularly in height, like the English contrast between [i] as in *heed* and [ɪ] as in *hid*, which are often very difficult for speakers of non-Germanic languages to distinguish. Many Germanic languages also contrast front rounded vowels from front unrounded vowels, such as [y:] versus [i:] in German (3).

The earliest inscriptions in a Germanic language date from the second century AD and are written in the runic script. The oldest runic alphabet is called the

elder futhark, a name formed by concatenating the sounds of the first six letters of that alphabet:

- (4) $\text{ƿ} \text{ ƒ } \text{ ƥ } \text{ ƚ } \text{ ʀ } \text{ ʁ}$
 f u θ a r k

The runic script was clearly inspired by the Latin script, but the origins of some of the letter forms are a mystery. All Germanic languages currently use the Latin script, but Icelandic still employs ⟨þ⟩, a version of runic ƥ, to spell [θ], a sound not found in Latin.

There are three major subbranches of Germanic, but one of these, East Germanic, has become extinct. It is best attested by a translation of the New Testament into Gothic. That translation dates from about 350 AD, centuries before extended texts were found in other Germanic languages. Consequently, Gothic words are usually favored by linguists delving into the history of Indo-European.

The two surviving branches are North Germanic and West Germanic. The modern North Germanic languages are all descendants of Old Norse, which was generally thought of as one language, albeit with some dialectal variation, until the 14th century. During the Viking period, Old Norse was extended to many Nordic colonies throughout Europe, including the Faroe Islands, Iceland, and much of Britain. Old Norse had a particularly strong influence on the development of English, which borrowed hundreds of Old Norse words beginning in the 9th century AD. These loans are mostly everyday words like *give* and *die* and even the pronoun *they*. North Germanic languages are notable for expressing the definite article as a suffix. The eastern languages – Danish, Norwegian, and Swedish – have a contrast between two different types of stress. In Norwegian and Swedish (5) the distinction is realized in the tone contours. In Danish, the distinction is manifested by the presence or absence of a glottal stop, or, more commonly, **CREAKY VOICE**. Also called *glottal fry*, this is a type of voicing in which the vocal folds vibrate very slowly and irregularly.

- (5) Swedish *anden*
 a. 'an\ɔd-ɛn\|
 duck-DEF.ART
 'the duck'
 b. 'an\ɔɛ-n\|
 spirit-DEF.ART
 'the spirit'

The West Germanic subbranch contains English, which is familiar to all readers of this book as well as to hundreds of millions of others who are native speakers

or use it as a second, international language (Crystal 2003). It has been very hospitable to loanwords from other languages, especially from Old Norse, Old French (FRO), and Latin. Despite occasional claims that such borrowing has turned English into a mixed language (§8.1), its basic character as a West Germanic language is unchanged. What is unusual about English is its spelling system. Due to its extremely conservative nature and propensity to spell loanwords as they were spelled in the donor language, there are many inconsistencies in the way it maps phonemes to letters. Unless Scots (SCO) is counted as a distinct language, the nearest relatives of English are the Frisian languages spoken primarily on islands off the coasts of Netherlands and Germany, although English speakers can understand very little of Frisian.

High Germanic refers to West Germanic languages and dialects spoken in the uplands of central Germany and further south. The literary standard of German is based on High Germanic. From a linguistic standpoint, High Germanic is noteworthy because it underwent a second consonant shift that was much like the earlier, pan-Germanic consonant shift. Voiceless stops became fricatives or affricates, and voiced stops became voiceless. Consider, for example, the word for 'eat'. Original [d], as seen in Latin *ed-*, became [t] in Germanic, by Grimm's law: thus English *eat*. Then, by the High Germanic shift, the [t] became the fricative [s], as seen in German *essen*. Yiddish also originated in the High Germanic region and shares in the High German shift: 'eat' is ײַסן [ɛsn]. Yiddish, whose name means 'Jewish', is written in the Hebrew script and has borrowed extensively from Hebrew and Slavic languages.

3.1.3 ITALIC

Around the fifth century BC, the Italic branch comprised a handful of languages spoken in central Italy. The best known of these were Oscan (OSC), Umbrian (XUM), Faliscan (XFA), and Latin (LAT). An important phonological marker of the Italic languages is that the Proto-Indo-European breathy stops, such as [b̥], became voiceless fricatives, such as [f]. Because the same sounds became voiced stops in Germanic by Grimm's law, the two branches have some rather dissimilar cognates, such as English *brother* corresponding to Latin *frater*.

Due to the ascendancy of Latin, which spread from Rome to the rest of Italy and on to a vast empire, the other Italic languages were supplanted and left no descendants. Therefore we shall focus here on the descendants of Latin, which are known as the Romance languages. Table 3.5 lists a selection of the Romance languages spoken today. The full list may be as much as three times larger, depending in part on whether certain varieties are considered independent languages or different dialects of the same language – a recurrent problem when making lists of languages. The large numbers of speakers of Romance languages reflect in part the imperial aspirations of the ancient Romans, who spread their language to much of Europe, and more recent imperial and colonial aspirations of the French, Spanish,

TABLE 3.5
Some Romance languages

Name	Size	Location
French (FRA)	9	France
Romansch (ROH)	5	Switzerland
Ladin (LLD)	5	Italy
Friulian (FUR)	6	Italy
Occitan (OCI)	7	France
Catalan (CAT)	8	Spain
Spanish (SPA)	9	Spain
Judeo-Spanish (LAD)	5	Turkey
Portuguese (POR)	9	Portugal
Italian (ITA)	8	Italy
Sardinian (SRD)	7	Italy
Romanian (RON)	8	Romania

and Portuguese nations. French is an official language in 29 countries and is widely learned elsewhere as a second language. Spanish is not far behind, at 20 countries. Many other Romance languages, by contrast, have shrunk dramatically, mostly because their speakers have shifted to speaking the dominant national language. Judeo-Spanish, which is spoken by Sephardic Jews, now mostly in Israel, is one of several minority Romance languages threatened with extinction. The language is perhaps more commonly called *Ladino*, but we stick with the more academic name to avoid confusion. The name Ladino is also used for another Romance language, Ladin (LLD), which is spoken in the Dolomite Mountains in Italy.

Romanian is totally surrounded by speakers of non-Romance languages and is a language of the descendants of Romans who colonized the Roman province of Dacia before the arrival of Slav and Magyar (Hungarian) tribes. Romanian has borrowed extensively from its Slavic neighbors.

3.1.4 GREEK

The Greek branch, which is often referred to as the Hellenic branch, consists of only one language, Greek. Ancient Greek (GRC), which goes back to the 9th century BC, has always been a very important source of information about early Indo-European. In 1952, Michael Ventris deciphered a collection of inscriptions found in Bronze Age sites in Greece, showing that they were written in a very old form of Greek, Mycenaean Greek (GMY), which pushed the attestation of Greek back to about 1400 BC. The script in which Mycenaean Greek was written, Linear B, was a syllabography with a leavening of logograms; its deciphering is described in Chadwick (1990). Greek is, along with Latin, one of the two major classical languages of European civilization and sources of international scientific vocabulary. It was also the main ecclesiastical language of Orthodox Christianity.

3.1.5 ALBANIAN

This branch, too, consists of a single language, Albanian, and its dialects. The language has borrowed extensively from Slavic, Greek, Latin, Italian, and Turkish. There are two main dialects: Tosk (ALS) in the south, which is the basis of the standard language, and Geg (ALN) in the north. The two dialects are said to be mutually unintelligible. The earliest inscriptions date from the sixteenth century AD.

3.1.6 BALTO-SLAVIC

The main Balto-Slavic languages are listed in Table 3.6.

3.1.6.1 Baltic

The Baltic subbranch is spoken in the Baltic countries of Lithuania and Latvia. The oldest Baltic document, a vocabulary list, dates from 1300 AD. It is written in Old Prussian (PRG), a language formerly spoken in East Prussia which became extinct in the early eighteenth century. Lithuanian is considered to be the most conservative of all the living Indo-European languages and is therefore extensively studied by Indo-Europeanists. Some dialects retain seven distinct cases in the noun system and have an accent system very much like that posited for Proto-Indo-European.

TABLE 3.6
Some Balto-Slavic languages

Name	Size	Location
Baltic	7	
· Lithuanian (LIT)	7	Lithuania
· Latvian (LAV)	7	Latvia
Slavic	9	
· Western	8	
· · Polish (POL)	8	Poland
· · Czech (CES)	8	Czechia
· · Slovak (SLK)	7	Slovakia
· · Sorbian (WEN)	5	Germany
· Southern	8	
· · Serbo-Croatian (HBS)	8	Serbia
· · Macedonian (MKD)	7	Macedonia
· · Slovenian (SLV)	7	Slovenia
· · Bulgarian (BUL)	7	Bulgaria
· Eastern	9	
· · Russian (RUS)	9	Russia
· · Ukrainian (UKR)	8	Ukraine
· · Belarusian (BEL)	8	Belarus

3.1.6.2 Slavic

The much larger subbranch of Balto-Slavic is Slavic. Slavic has split into three main subgroups, but a great deal of uniformity obtains among all the Slavic languages because the split did not occur until around the 11th century AD. Old Church Slavonic (CHU) was established as a literary and liturgical language in the 9th century and so provides excellent information about Slavic prior to its breakup.

The West Slavic group includes national languages in Poland, the Czech Republic, and Slovakia; the languages of the latter two countries are particularly closely related to each other. In our illustrations we refer to the Czech Republic as *Czechia*, a synonym that is shorter if not as well known to English speakers. Sorbian is a Slavic pocket located within Germany, near Poland and the Czech Republic. It should not be confused with Serbian, which is a South Slavic language form. The West Slavic territory is historically Roman Catholic, and consequently the writing systems of these languages use the Latin script.

The South Slavic languages are spoken in the Balkans and are separated from the other Slavic languages by Austria, Hungary, and Romania. All the South Slavic languages form a DIALECT CONTINUUM, which means that the language spoken in one village is very similar to the language spoken in the next village. The language varieties at each end of the continuum may be quite different from each other, but there is no linguistic dividing line between the languages. One common way of dividing up languages is to draw the lines at national borders, which is often expedient even though it corresponds to no real differences in the languages themselves.

Macedonian and Bulgarian are especially close to each other linguistically, and both have lost virtually all case inflections. Serbo-Croatian, on the other hand, is much more conservative, preserving seven cases inherited from Proto-Indo-European. Along with Slovenian, it has a pitch accent, such that words can be distinguished by whether the vowel has a rising, falling, or neutral pitch. Serbo-Croatian is the majority language in Bosnia and Herzegovina, Croatia, Montenegro, and Serbia, violating the expectation that language names should correspond with national boundaries. Many people feel this discrepancy particularly strongly because of the emphasis on establishing separate national identities after the violent dissolution of Yugoslavia. Consequently, one often hears of the Serbian (SRP) language, the Croatian (HRV) language, and so forth. Linguistically, however, these languages are very nearly identical. The linguist's traditional hyphenated name, *Serbo-Croatian*, follows a common idiom and means 'the language spoken from Serbia to Croatia'. Diplomats, who apparently have more time on their hands, sometimes call this language *Bosnian/Croatian/Montenegrin/Serbian*, listing all the countries, in alphabetical order. South Slavic languages vary in whether they use the Latin or the Cyrillic alphabet. For the most part, the choice depends on whether the region traditionally worshipped according to western (Latin) or eastern (Greek and Cyrillic) rites. But in Eastern Orthodox Serbia, where the Cyrillic is the official script, the Latin script is very popular.

The East Slavic languages are contiguous to the West Slavic ones and use the Cyrillic script. Russian has by far the most speakers of any Balto-Slavic language and will be described in detail in a sketch at the end of this chapter (§3.5).

3.1.7 ARMENIAN

Armenian is another branch of Indo-European that consists of a single language. The vocabulary of Armenian contains many Iranian loanwords, which led early scholars to believe it was an Iranian language. The oldest inscriptions date from the fifth century AD. An Armenian monk devised a special alphabet for the language which is in use to the present day. In that alphabet, the word for 'Armenian' is Հայերեն [haje'ren].

3.1.8 INDO-IRANIAN

This and the following branches of Indo-European are located primarily in Asia. Close to a billion people speak languages belonging to the Indo-Iranian branch of Indo-European. It consists of two subbranches, as listed in Table 3.7.

TABLE 3.7

Some Indo-Iranian languages

Name	Size	Location
Indic	9	
· Sanskrit (SAN)	5	India
· Hindi (HIN)	9	India
· Punjabi (PAN)	9	Pakistan
· Gujarati (GUJ)	8	India
· Marathi (MAR)	8	India
· Bengali (BEN)	9	Bangladesh
· Nepali (NEP)	8	Nepal
· Sinhala (SIN)	8	Sri Lanka
· Urdu (URD)	9	Pakistan
· Kashmiri (KAS)	7	India
· Romani (ROM)	7	Romania
Iranian	9	
· Persian (FAS)	8	Iran
· Tajik (TGK)	7	Tajikistan
· Pashto (PUS)	8	Afghanistan
· Balochi (BAL)	8	Pakistan
· Kurdish (KUR)	8	Turkey
· Ossetic (OSS)	6	Russia
· Old Persian (PEO)	0	Iran
· Avestan (AVE)	0	Afghanistan

The Indic group includes most of the languages spoken in the northern parts of the Indian subcontinent. The national languages of most of the countries in the region are Indic: Hindi in India, Urdu in Pakistan, Bengali in Bangladesh, Nepali in Nepal, Sinhala in Sri Lanka, and Dhivehi (DIV) in the Maldives. Hindi and Urdu are extremely similar to each other, but are differentiated along religious lines, with Hindi being spoken by Hindus and Urdu by Muslims. Hindi and Urdu are nowadays written in different scripts – Devanagari and Arabic, respectively – and prefer to draw on different sources for loanwords. Hindi draws largely from Sanskrit, and Urdu draws largely from Persian and Arabic.

Most of the Indic languages, like Hindi, are written with abugidas that are derived from the Brahmi script. See, for example, our discussion of the Gujarati abugida in §2.1.3.4. As Table 2.3 shows, this family of scripts is by no means limited to the Indic languages, but has also been adapted for many unrelated languages in greater India and points east.

Included among the Indic languages is Romani, even though it is spoken primarily in Europe, not the Indian subcontinent. The latest hypothesis concerning the origin of the Roma people is that they left India sometime around 1000 AD, eventually reaching various parts of Europe by different routes. In English, the Roma are more commonly known as *Gypsies*, but the current trend has been to avoid that name. Not only is it sometimes used offensively, but it is also ludicrously inaccurate, being a shortening of the word *Egyptians*. Romani dialects have borrowed heavily from the languages of the regions where they were spoken, but they retain Indic basic vocabulary and inflections; this is especially true of Romani as spoken in southeastern Europe. In other regions, such as the United Kingdom and the Iberian peninsula, Roma use the local language, often augmented with many Romani words.

The earliest records in an Indic language are the orally transmitted religious hymns in Vedic Sanskrit, which may have been composed as early as 1400 BC but were not written down until more than 1,000 years later, when writing came to India. Sanskrit began evolving into the Indic languages by 300 BC, but for an ancient language it still enjoys a great deal of vitality in India. In the 2001 Indian census, over 14,000 people reported that Sanskrit was their mother tongue (Government of India 2006).

The Iranian group is spoken in regions northwest of the Indic group. Iranian languages are the predominant languages from Iran to Tajikistan, including Afghanistan and much of Pakistan. Kurdish is a noteworthy case of a language not constrained by modern national borders, being spoken in Iran, Iraq, Syria, and Turkey, among other countries. Persian likewise is spoken across several countries, with opinions varying as to whether local forms such as Tajik in Tajikistan and Dari (PRS) in Afghanistan should be considered dialects of Persian or separate national languages. Most Iranian languages are spoken predominantly by Muslims, are traditionally written in Arabic script, and often draw vocabulary from the great Persian literary tradition. An earlier religion of ancient Persia was

Zoroastrianism, whose primary religious texts were composed in Avestan, which furnishes important information about early stages of Iranian languages. Like the Sanskrit Vedas, they were transmitted orally for centuries before being committed to writing. The oldest inscriptions in any Iranian language are Old Persian and date to 550 BC.

3.1.9 ANATOLIAN

The major representative of the Anatolian branch of Indo-European is Hittite. Inscriptions in this language date back to 1400 BC and are the earliest known inscriptions in any Indo-European language. This branch has been extinct for a long time. Originally consisting of several languages spoken in Asia Minor, in what is now Turkey, it was discovered only in the 20th century. The discovery of this branch caused Indo-Europeanists to reconsider their theories concerning the nature and development of Proto-Indo-European. In many respects the Anatolian languages were quite different from other ancient Indo-European languages such as Sanskrit and Ancient Greek. For example, Hittite distinguishes singular from plural but has no special inflections for the DUAL NUMBER – a separate grammatical category for referring to two things. Nor does it have the masculine–feminine–neuter gender system found in many Indo-European languages like Russian (§3.5.4.1). Innovations within Anatolian could account for many of these differences, but many linguists believe that some of the peculiarities of Anatolian are inherited retentions (Melchert, in press). This point of view implies that Anatolian may have been the first clade to split away from Proto-Indo-European. The rest of the Indo-European community may have then experienced several shared innovations before they split up.

3.1.10 TOCHARIAN

Tocharian, too, is an extinct branch of Indo-European. It was represented by two closely related languages spoken in several oases in the Tarim Basin of Central Asia, which is now the Xinjiāng Uyghur Autonomous Region of China. Texts date from around the seventh century AD and are mostly Buddhist religious writings. These languages may have been the source of some very early loanwords in Chinese. For example, the Old Chinese (OCH) word for ‘honey’ was [mjit], which is very similar to the Tocharian B word for the same concept, [mit] (Polivanov 1916). The latter is related through Proto-Indo-European to the English *mead* [mid], a fermented honey drink.

The discovery of Tocharian early in the 20th century helped to shake up some earlier ideas about the subgrouping of the Indo-European languages. Proto-Indo-European is reconstructed as having three different places of articulation for velar stops: one pronounced more forward in the mouth [k̟], one more back in the mouth [k̠], and one like the latter but also with lip rounding [k̠ʷ].

Table 3.8 presents a set of corresponding words for the more forward voiceless velar, [k̥].

Table 3.9 summarizes the correspondences for all three of the voiceless Proto-Indo-European velars across the different branches of Indo-European. The data for other types of velars are analogous. An intriguing pattern can be discerned in this table. In four of the branches, the front velar has become identical with the already existing back velar. In the other three branches, the front velar has become a SIBILANT CONSONANT. Sibilants are fricatives or affricates that make a particularly loud and high-pitched sound by forcing a stream of air at the teeth. In the same branches in which the front velar became a sibilant, the labialized velars lost their lip rounding, becoming identical with the back velar. This impressive congruity led 19th-century linguists to conclude that Proto-Indo-European split into two subgroups, which were characterized by those sound changes. They were called the *centum* and *satem* subgroups, respectively, after representative words for ‘hundred’ in those two groups (see again Table 3.8).

The idea that Proto-Indo-European split into two groups, each with its own unique sound changes for velar stops, was reinforced by the observation that the

TABLE 3.8

Development of */k̥/ in Indo-European languages

Branch	Language	[k̥m̥tóm] ‘hundred’
Italic	Latin	[k̥entum] <i>centum</i>
Greek	Greek	[he-k̥atón] ἑκατόν
Celtic	Welsh	[k̥ant] <i>cant</i>
Germanic	Gothic	[h̥und] 𐌺𐌹𐌸𐌰
Indo-Iranian	Avestan	[s̥atəm] 𐬀𐬎𐬀𐬎𐬀
Indo-Iranian	Sanskrit	[s̥atə] शत
Balto-Slavic	Russian	[st̥o] сто
Balto-Slavic	Lithuanian	[š̥mtas] šimtas
Armenian	Armenian	[s̥i:t] սիրտ < PIE [k̥e:rd] ‘heart’

TABLE 3.9

Basic correspondences of voiceless velar stops in the branches of Indo-European

Branch	*k̥	*k	*kʷ
Italic	k	k	kʷ
Greek	k	k	kʷ
Celtic	k	k	kʷ
Germanic	x	x	xʷ
Indo-Iranian	ç	k	k
Balto-Slavic	ç	k	k
Armenian	s	k	k

centum languages cluster together in the west and the satem languages cluster together in the east. However, Tocharian, the easternmost Indo-European branch, turned out to be a centum language, changing *[k̥] to [k] rather than to a sibilant. Suddenly, with thousands of miles of satem languages separating Tocharian from the nearest centum language, it no longer seemed as plausible that all the centum languages form a coherent subgroup.

Nowadays few linguists believe that Proto-Indo-European split directly into a centum clade and a satem clade, and the isolated position of Tocharian is not even the most important reason for the skepticism. If that really were an old and important split within Indo-European, we would expect the centum languages to share many other innovations that the satem languages lack, and vice versa. That isn't the case. The satem innovations, like many others encountered in two or more branches of Indo-European, are believed to have spread from one branch to another after the branches had split from the parent language.

3.1.1.1 GENERAL FEATURES OF INDO-EUROPEAN

Proto-Indo-European was a highly inflected, fusional language. Languages that conserve much of that morphology include several of the ancient and classical languages such as Sanskrit, Greek, Latin, and Old Church Slavonic; even today, most of the Balto-Slavic languages have a wealth of inflectional affixes. The conservative languages often retain a half dozen different case suffixes on nouns; three grammatical genders (feminine, masculine, and neuter); separate inflections for singular, plural, and sometimes dual number; verbs that inflect for person, number, tense, and mood; and markers of agreement between nouns and their modifiers and between verbs and their subject. Many of these features will be explained in the sketch of Russian (§3.5). But many more-innovating languages, such as English and French, have reduced or eliminated case suffixes, relying more heavily on word order and prepositions to show the grammatical relations between words. Gender distinctions are sometimes reduced to two categories or are lost altogether.

Most modern Indo-European languages have SVO word order. The most notable exceptions are the Indo-Iranian languages, which tend to be SOV, and the modern Celtic languages, which are VSO. Indo-European languages have nominative-accusative alignment. Again, most of the exceptions are to be found among the Indo-Iranian languages, many of which use ergative constructions in at least some contexts.

3.2 Uralic

Like Indo-European, the Uralic family extends from Europe to Asia (Figure 3.2). The national languages that belong to this family are Finnish, Estonian, and Hungarian.



FIGURE 3.2 Some Uralic languages

TABLE 3.10

Some Uralic languages

Name	Size	Location
Samoyedic	5	Russia
· Enets (ENH)	2	Russia
· Nenets (YRK)	5	Russia
· Nganasan (NIO)	3	Russia
· Selkup (SEL)	4	Russia
Sami	5	
· Northern Sami (SME)	5	Norway
· Southern Sami (SMA)	3	Sweden
Permic	6	Russia
· Udmurt (UDM)	6	Russia
· Komi (KOM)	6	Russia
Mordvinic	6	Russia
· Erzya (MYV)	6	Russia
· Moksha (MDF)	6	Russia
Finnic	7	
· Finnish (FIN)	7	Finland
· Estonian (EST)	7	Estonia
· Karelian (KRL)	5	Russia
Khanty (KCA)	4	Russia
Mari (CHM)	6	Russia
Hungarian (HUN)	8	Hungary

The rest are less populous languages spoken in the northern regions of Europe and Asia. Table 3.10 lists the main living Uralic languages.

3.2.1 SAMOYEDIC

The Samoyedic branch of Uralic consists of a handful of endangered languages spoken in Siberia, which are grouped under the names Nganasan, Enets, Nenets,

and Selkup. The first two are spoken in the Taymyr Peninsula, the northernmost part of the Eurasian mainland. The name of the branch links its speakers with the Sami, another group of Uralic speakers who live in the far north.

The Samoyedic languages were the last major group of languages to be identified as being in the Uralic family. There is a tendency to think of all the Uralic languages that had been identified previously as composing a separate, Finno-Ugric, clade within Uralic. But the comparative method lends only weak support to the idea that they constitute a separate genetic clade (Salminen 2002). Therefore we describe smaller subgroups in §3.2.2–§3.2.6 rather than a hypothesized Finno-Ugric group.

3.2.2 SAMI

The Sami languages are spoken in northern Europe: Sweden, Norway, Finland, and Russia. All Sami languages are considered endangered (Moseley 2010), but the least endangered is Northern Sami, which is spoken in the very northernmost reaches of the three Scandinavian countries. Its older name, *Lapp*, familiar from the English name *Lapland*, is now considered insulting. That is often the fate of EXONYMS, names bestowed by speakers of other languages.

3.2.3 PERMIC

The Permic languages are Udmurt and Komi. They are spoken in the western foothills of the Ural mountains.

3.2.4 MORDVINIC

The Mordvinic group embraces two languages spoken mostly in Mordovia in Russia: Moksha and Erzya.

3.2.5 FINNIC

The Finnic branch includes the national languages Finnish and Estonian. Finnish exemplifies many of the characteristic features of the Uralic languages, which will be taken up in this chapter's sketch of Finnish (§3.6). Estonian has attracted much attention from phonologists because it appears to contrast three degrees of phonemic length – short, long, and overlong; other languages have two degrees at most. For example, [vina] means 'vapor', [vi:na] means 'vodka' in the genitive case, and [vi::na] means 'vodka' in the PARTITIVE CASE ('some of the vodka'). Karelian is spoken mostly in the Karelian Republic in Russia. An oath inscribed on birch bark in the 13th century AD is the oldest known Finnic inscription.

3.2.6 OTHER URALIC LANGUAGES

In Russia, Mari is spoken in the Mari Republic, and Khanty and Mansi (MNS) are spoken in the Khanty-Mansi Autonomous Okrug.

Hungarian, whose native name is *Magyar*, is spoken in Hungary and neighboring countries, including Romania, Serbia, and Slovakia. The ancestors of the modern Hungarians arrived in present-day Hungary around 900 AD after a series of migrations from a region just west of the Urals as a part of a military coalition of Magyar and various Turkic tribes. The earliest inscriptions in Hungarian date to about 1200 AD.

3.2.7 GENERAL FEATURES OF URALIC

Some linguists feel that there is enough evidence to link this family with Indo-European. But borrowing may be a simpler interpretation of this evidence. Uralic speakers have long been in contact with Indo-European speakers – it may even turn out, as some scholars have speculated, that Proto-Indo-European and Proto-Uralic were neighboring languages. There has clearly been steady borrowing from Baltic, Iranian, Slavic, and Germanic languages. In fact, Finnish loanwords from Germanic have preserved some very archaic Germanic features. Thus, although genetic relationship cannot be ruled out between Indo-European and Uralic, it is at present difficult, if not impossible, to ascertain which similarities are due to genetic relationship and which to DIFFUSION (borrowing) of linguistic traits across language family boundaries.

Typologically, Uralic languages range from agglutinative (e.g. Finnish) to fusional (e.g. Estonian). According to Janhunen (1992: 208), Proto-Uralic was agglutinative and had SOV order. Several languages have well over a dozen cases.

There is usually fairly complex verb morphology in Uralic languages. One of the most striking features is that many Uralic languages have special negative verbs that are used as auxiliary verbs to negate other verbs. Compare Finnish [en ota] ‘I don’t take’ to [em:e ota] ‘we don’t take’. The stem of the negative verb is [e-], and the personal suffixes [-n] and [-mme] are attached to it.

There are some typologically rare phenomena in Uralic languages. A number of them, but not all, exhibit CONSONANT GRADATION. Changes take place before the vowel of a closed syllable: long stops become short, and short stops become voiced. For example, in Finnish, [ak:a] ‘old woman’ becomes [akat] when the plural suffix [-t] is added, causing the final syllable to be closed. In the same situation, [pata] ‘cooking pot’ becomes [padat] (see further §3.6.2.3 and Collinder 1965: 67–73).

Most Uralic languages also exhibit VOWEL HARMONY: vowels within the same word must be similar to each other. Vowel harmony differs from language to language with respect to its domain. In some languages, vowels may harmonize only within a morpheme; in others, they may harmonize between a root and all its suffixes, or even across all the elements of a compound word. In Finnish, vowel harmony extends throughout a root and its suffixes, but a root in a compound word can have vowels that do not harmonize with the vowels in the other root of the compound. Vowel harmony also varies with respect to which vowel features are

required to harmonize. Most typical of Uralic is FRONT-BACK HARMONY, as found in Finnish (3.6.2.2), in which certain front vowels cannot occur in the same word as back vowels. Hungarian and Meadow Mari have in addition ROUNDEDNESS HARMONY. Vowel harmony is very prominent in Uralic because of the prevalence of suffixes. Many suffixes come in two or more different forms so that the speaker may choose the variant that does not violate the vowel harmony rules.

3.3 Caucasus area

In the Caucasus region between the Black Sea and the Caspian Sea are spoken about three dozen languages that are members of three language families spoken nowhere else. A selection of these languages is listed in Table 3.11 and located in Figure 3.3.

TABLE 3.11
Some languages of the Caucasus

Name	Size	Location
West Caucasian	7	Russia
· Abkhaz-Abaza	6	Georgia
· · Abkhaz (ABK)	6	Georgia
· · Abaza (ABQ)	5	Russia
· Circassian	7	Russia
· · Kabardian (KBD)	7	Russia
· · Adyghe (ADY)	6	Russia
· Ubykh (UBY)	0	Turkey
East Caucasian	7	Russia
· Nakh	7	Russia
· · Chechen (CHE)	7	Russia
· · Ingush (INH)	6	Russia
· · Batsbi (BBL)	4	Georgia
· Daghestanian	7	Russia
· · Avar (AVA)	6	Russia
· · Andi (ANI)	4	Russia
· · Tsez (DDO)	5	Russia
· · Lak (LBE)	6	Russia
· · Dargwa (DAR)	6	Russia
· · Lezgi (LEZ)	6	Russia
· · Tabassaran (TAB)	6	Russia
Kartvelian	7	Georgia
· Georgian (KAT)	7	Georgia
· Svan (SVA)	5	Georgia
· Laz (LZZ)	6	Turkey



FIGURE 3.3 Some languages of the Caucasus. ¹West Caucasian. ²East Caucasian. ³Kartvelian.

3.3.1 WEST CAUCASIAN

The languages of the West Caucasian family are spoken mainly in the northwestern part of the Caucasus region, but some of them are also spoken in scattered pockets elsewhere. Many people from this region resettled in the Ottoman Empire, especially modern Turkey, after the Russian conquest of 1864, and tens of thousands of emigrants still speak their West Caucasian languages. Abkhaz is spoken mainly in the Republic of Abkhazia in Georgia. The others are spoken primarily in several republics of Russia, where they are official languages: in Karachay-Cherkessia, in Adygea, and in Kabardino-Balkaria and Karachay-Cherkessia. The last speaker of Ubykh died in 1992.

Typologically, these languages have ergative morphosyntactic alignment, lack the distinction between active and passive voice, and have SOV word order. They have complex verb morphology but relatively simple noun morphology. The verbs agree in gender with the subject, object, and indirect object. The languages are generally agglutinative in structure, with prefixes outnumbering suffixes.

One of the noteworthy features of these languages is that they possess very rich consonant phoneme inventories but very few vowel phonemes. Thus, for example, Ubykh has been analyzed as having as many as 82 consonant phonemes. This number is due to consonants that vary with respect to features like lip rounding, palatalization (discussed further in the sketch of Russian §3.5), and glottalization (with full or partial closure of the glottis). In fact, Ubykh was once cited in the Guinness Book of World Records for having the largest consonant inventory in the world (Young 1996: 128). But it was analyzed as having only two vowel phonemes.

3.3.2 EAST CAUCASIAN

The East Caucasian language family consists of two branches (Nichols 2013).

In the Nakh branch, Chechen and Ingush are spoken primarily in the Russian republics of Chechnya and Ingushetia. Batsbi is spoken just across the border in a village in Georgia.

The Daghestanian branch is spoken mainly in the Russian Republic of Dagestan. It consists of numerous languages of various sizes, of which only a selection is listed in Table 3.11. Nichols 2013 recognized six subgroups: Avar-Andic, which includes Avar and Andi; Tsezic, which includes Tsez; Lak; Dargwa; Lezgian, which includes Lezgi and Tabassaran; and Xinalug (кӀӀ).

Typologically, the East Caucasian languages are ergative, although they do show some nominative-accusative traits. Their basic word order is SOV. As opposed to the West Caucasian languages, their verb morphology is very simple, whereas their noun morphology is very complex. It is often claimed that Tabassaran has the world's record for number of cases, 48 (Young 1996: 128), though such a large number may be misleading. It has about 14 or 15 case suffixes, most of which express location such as 'in', 'on', and 'behind'. Many of these suffixes can be regularly combined with a suffix meaning 'to' or one meaning 'from' to get combinations such as 'into' or 'from behind'. The number 48 comes from adding up all the attested combinations of that appreciably smaller set of basic case endings (Comrie & Polinsky 1998).

3.3.3 KARTVELIAN

The most prominent Kartvelian language is Georgian, which has the largest number of speakers of any Caucasian language and is the official language of Georgia. In this group of languages both SVO and SOV word orders are found, sometimes in the same language.

Georgian behaves in some respects like an ergative-absolutive language and in other respects like a nominative-accusative one. In the imperfective aspect (§3.5.4.5), the subject of a transitive verb takes the same case as the subject of an intransitive verb (6a; these examples are from Hewitt 1987), and the object of the transitive verb takes a special case (6b), just as in nominative-accusative languages.

(6) Georgian imperfective

- a. k'ats-i kal-s k'lavs
 man-NOM woman-DAT kills

'The man kills the woman.'

DAT = DATIVE CASE, which is more typically used for indirect objects of verbs like 'give'.

- b. k'ats-i k'vdeba
man-NOM dies

'The man dies.' (both examples from Hewitt 1987)

But in the perfective aspect, the subject of a transitive verb takes a special case, the ergative (7a), while the subject of an intransitive verb takes the same case as the object of a transitive verb (7b), just as in ergative-absolutive languages.

(7) Georgian perfective

- a. k'ats-ma kal-i mok'la
man-ERG woman-NOM killed

'The man killed the woman.'

- b. k'ats-i mok'vda
man-NOM died

'The man died.' (both examples from Hewitt 1987)

Languages that are ergative-absolutive only in some types of constructions are said to have *SPLIT ERGATIVITY*. It is this split ergativity that is responsible for some of the confusing case names in Georgian. If all sentences were ergative-absolutive, what we have marked as *nominative* in these last two sentences would be called *absolutive*.

To make matters a little more complicated, Georgian has some intransitive verbs whose subject takes the ergative. Confusingly, the verbs that take the ergative case marking are called *UNERGATIVE*, because they behave in a way that intransitive verbs are not supposed to behave in an ergative-absolutive language. Languages whose intransitive verbs vary in such a way are called *SPLIT-S* languages, using a convention whereby S stands for the subject of an intransitive verb. In split-S languages, the intransitive verbs that take ergative subjects often appear to be those whose subject is the source or originator of the action or has some volitional control over the action. The precise semantics is hard to pin down, however, because individual verbs in Georgian are locked into using a specific case without considering whether the action being described was volitional on that specific occasion.

3.3.4 GENERAL FEATURES OF CAUCASUS LANGUAGES

The majority of Caucasus languages share a number of traits. Most of them are ergative-absolutive to some degree, have subject and object agreement affixes on the verb, and are generally agglutinative in morphology. They also have *PHARYNGEAL CONSONANTS*, and *EJECTIVE CONSONANTS*, the latter being a type of glottalized sound (the larynx is pushed up to force air through the mouth). Unless it is just a coincidence that these languages share such rather rare features, there

are two possibilities. They could be AREAL FEATURES, features that have diffused through the Caucasus. Under this hypothesis, the Caucasus is a SPRACHBUND, a linguistic area whose languages are quite similar due to massive borrowing. The second possibility is that the languages inherited these features from a common ancestor. Writing against the first possibility, Klimov (1965: 77) pointed out that the Caucasus region is, geographically speaking, a very unlikely candidate to be a linguistic area: the mountainous terrain greatly hinders communication and contact between settlements. On the other hand, all Caucasus languages show large numbers of loanwords from Iranian languages, Turkic languages, and Arabic. Presumably, therefore, the mountainous terrain does not entirely prevent language contact. Because the comparative method does not yield convincing evidence that the languages are related, the sprachbund hypothesis appears more plausible.

3.4 Basque

Basque is an ISOLATE: a language with no known relatives. It is spoken by a fairly small number of people in the western Pyrenees, some in France but most in Spain. Although it has a relatively free word order – that is, the main components of a verb phrase can appear in virtually any order – its neutral word order appears to be SOV. Adjectives and nouns in the genitive case precede the nouns they modify. Its morphology is agglutinative, with 16 cases. Basque is commonly referred to as an ergative-absolutive language, but it is more properly described as split-S, because some intransitive verbs take subjects that have the same case as subjects of transitive verbs. Compare the examples in (9) in Chapter 1 with (8) here, where the unergative verb meaning ‘resign’ takes the ergative case:

(8) Basque

presidente-a-k dimititu du
 president-DEF-ERG resign AUX

‘The president has resigned.’ (Hualde & Ortiz de Urbina 2003: 181)

Basque has been influenced by the languages it has had contact with: Latin, Spanish, French, and the Gascon dialect of Occitan, which is spoken in southwest France. For example, the Zuberoan dialect of Basque, which is spoken in Soule, France, has the vowel [y], which is lacking in the Basque dialects spoken in Spain. Its existence in Zuberoan may have been encouraged by the fact that Occitan and French have that vowel, but Spanish does not.

Amateur historical linguists have given Basque an unusual amount of attention, constantly attempting to connect it genetically with other languages. Larry Trask, who was a leading expert on Basque, felt compelled to plead on his Web page: “Please note: I do not want to hear about . . . your latest proof that Basque is related to Iberian/Etruscan/Pictish/Sumerian/Minoan/Tibetan/

Isthmus Zapotec/Martian” (A. Brown 2003). Part of the reason that Basque fascinates the public is that it is a language isolate. But there is nothing genetically special about a language isolate: language families may have any number of members, from one, two, three, up to several hundred. Moreover, the designation *isolate* hinges crucially on the quite arbitrary question of whether related language varieties are to be considered different languages or different dialects. If the five main dialects of Basque, which are noticeably different from each other, were listed as different languages, then Basque would not be considered an isolate. Amateur attempts to connect Basque to other languages often focus on ergative-absolutive languages, including those of the Caucasus and Burushaski, another language isolate, spoken in Pakistan (§4.8). None of these attempts has won widespread support among linguists.

3.5 Sketch of Russian

3.5.1 GENERAL BACKGROUND

Russian belongs to the East clade of the Slavic branch of the Indo-European language family. It is spoken by about 270 million people, including 150 million who consider it their native language, primarily in Russia. The Soviet Union contributed greatly to the spread of Russian both by facilitating migration of ethnic Russians to neighboring countries and by mandating instruction in Russian as the language of interethnic communication throughout the union. Since the breakup of the union in 1991, the former member states increasingly emphasize their national languages, but Russian is still an official language in Belarus, Kazakhstan, and Kyrgyzstan. It is also widely spoken in many other places with large ethnic Russian minority populations, including Ukraine, Latvia, Moldova, Estonia, and Georgia. However, Russian was never as widely taught or spoken outside the Soviet Union as one would expect given the political and military status of the union. Since Slavic languages are to a large extent mutually intelligible, many Slavs did not bother learning Russian because they felt that they could communicate adequately with Russians by speaking their own native language. In addition, not learning Russian or learning it badly was one of the ways in which people in the Eastern Bloc countries showed opposition to their domination by the Soviet Union.

Russian is a fusional language with a fairly complicated morphology and phonology. Its basic word order is SVO, and it places adjectives before nouns and head nouns before relative clauses.

Russian has borrowed extensively from other languages. In the past it was influenced by various Turkic languages, Uralic languages, Greek, Latin, and Germanic languages. In more recent times a large number of terms, especially technical terms, were borrowed from German, Dutch, French, and even more recently, English.

3.5.2 ALPHABET

The alphabet used to write Russian is called *Cyrillic*, in honor of Cyril, a ninth-century Greek missionary who was instrumental in spreading Christianity among the Slavs. Most of the letters of the Cyrillic alphabet were derived from the Greek alphabet; letters for a few sounds that did not have Greek equivalents were borrowed from Glagolitic, another early script for writing Slavic languages. Traditionally, the Cyrillic alphabet has been used for languages whose speakers predominantly belong to the Orthodox Church: Bulgarian, Serbian, Macedonian, Ukrainian, and Belarusian, in addition to Russian. During Soviet times, dozens of minority languages were furnished with Cyrillic writing systems; these included Uzbek, Kazakh, Tajik, and Azerbaijani. With the break-up of the Soviet Union, several newly independent countries moved toward adopting other writing systems, such as Latin in Azerbaijan.

3.5.3 PHONETICS AND PHONOLOGY

The consonantal phoneme inventory of Russian is shown in Table 3.12. Russian has an unusually large number of sibilants. As a so-called satem Indo-European language, it has sibilants in place of the Proto-Indo-European fronted velars; for example, [z] in [zna-] ‘know’ from [gnoh₃-] (9a; §3.1.10). Along with other Slavic languages, it has also undergone many other changes that turn sounds into sibilants. These changes all consisted of ASSIMILATING consonants to adjacent [j] or front vowels like [i] and [e]. Assimilation is any sound change that makes a sound more like another nearby sound. In this case, the assimilations involved place of articulation: the affected consonants were drawn toward the front of the hard palate where the conditioning sounds are produced. Such changes are called PALATALIZATIONS and are very common in the languages of the world. The outcomes of palatalizations can be quite varied, even within the same language. (9b–d) list the main outcomes of several palatalizations in the history of Russian that produced sibilants. The most recent palatalization (9e) is comparatively mild, in that it hasn’t changed the main place of articulation of the consonants very much, but it has

TABLE 3.12

Consonant phonemes of Russian

p ⟨п⟩	pʲ	t ⟨т⟩	tʲ			k ⟨к⟩
b ⟨б⟩	bʲ	d ⟨д⟩	dʲ			g ⟨г⟩
		ʃ ⟨ш⟩		ʃ̟ ⟨ш̟⟩		
f ⟨ф⟩	fʲ	s ⟨с⟩	sʲ	ʃ ⟨ш⟩	ʃ̟ ⟨ш̟⟩	x ⟨х⟩
v ⟨в⟩	vʲ	z ⟨з⟩	zʲ	ʒ ⟨ж⟩	ʒ̟ ⟨ж̟⟩	
m ⟨м⟩	mʲ	n ⟨н⟩	nʲ			
			rʲ	r ⟨р⟩		
		ʎ ⟨й⟩	ʎʲ			j ⟨й̟⟩

added a secondary articulation to their pronunciation: the dorsum of the tongue is raised toward the front of the hard palate. Thus [pʲ] is like pronouncing a [p] and a [j] simultaneously.

- (9) a. Unconditioned satem change: [k̚] > [s]; [g̚] > [z]; [g̚] > [z]
- b. Before [j], [i], or [e]: [k̚] > [t͡ɕ]; [g̚] > [ʒ]; [x̚] > [ʃ]; [sk̚] > [ç];
- c. Before [æ] or after [i]: [k̚] > [t͡s]; [g̚] > [z]; [x̚] > [s]
- d. Fusion: [tj̚] > [t͡ɕ]; [tsj̚] > [t͡ɕ]; [dj̚] > [ʒ]; [sj̚] > [ʃ]; [stj̚] > [ç]; [zj̚] > [ʒ]
- e. Before [j], [i], or [e]: [p̚] > [pʲ]; [t̚] > [tʲ]; [d̚] > [dʲ]; [s̚] > [sʲ]; [r̚] > [rʲ];
etc.

A pervasive distinction is drawn between flat and sharp consonants. These terms are attempts at characterizing an acoustic difference between the sounds in each class. Sharp consonants sound lighter, thinner, or noisier than flat consonants; Russian grammarians traditionally call them *soft*. Sharp consonants have a higher pitch component than corresponding flat consonants, because they are pronounced with the middle part of the tongue at the hard palate, a gesture like that produced when making an [i] or [j], which have especially high second formants (Bondarko 2005). More precisely, many sharp consonants are formed by adding the secondary articulation of palatalization to the articulation of the flat consonant; this is indicated by the IPA diacritic [ʲ]. Sometimes this produces side effects: sharp [tʲ] and [dʲ] are affricated, and [rʲ] sounds like a fricative. The alveolopalatal sounds [t͡ɕ], [ç], and [ʒ], as well as the palatal glide [j], are also considered sharp. The alveolopalatal sounds can be contrasted with the flat sounds [t͡s], [ʃ], and [ʒ].

Although we discussed the various palatalizations in historical terms, they continue to play an important role in Russian morphophonemics. When adding suffixes to a stem, the final consonant of the stem often changes between forms that were historically the inputs and outputs of palatalizations. For example, [pʲi'k-u] means 'I bake', but [pʲi't͡ɕ-ɔʃ] means 'you SG bake'. As this example shows, the front vowel or [j] that was historically responsible for the assimilation is very often no longer present; therefore, there may be no phonological clue as to which consonant to use. Consequently, sounds like [t͡ɕ] are no longer allophones of sounds like [k̚] but are now separate phonemes.

When an obstruent appears at the end of a word, it becomes voiceless; for example, in (10a), the end of the stem [god] DEVOICES when it has no suffix. Russian also has a rule of anticipatory, or *regressive*, voice assimilation across obstruents. An anticipatory assimilation occurs when a sound becomes more like a sound that is coming up. When two obstruents come together in Russian, the first one takes on the voicing of the latter one (10b). This even happens across word boundaries. If more than two obstruents come together, the assimilation keeps applying iteratively: the second takes on the voicing of the third, then the first takes on the new voicing of the second.

- (10) a. 'god-i → 'got
 year-M.NOM.PL year[M.NOM.SG]
 'years' 'year'
- b. 'gos'tʃ → 'goz'dʃ da'vid-a
 guest[M.NOM.SG] guest[M.NOM.SG] David-M.GEN.SG
 'guest' 'David's guest'

Russian orthography represents the etymology, or the underlying pronunciations, rather than the results of devoicing and voicing assimilation. For example, the word [got] 'year' is spelled год, ending with the letter that normally spells [d] (11a). The spelling represents the underlying phonological representation as shown by forms of the word where the [d] is followed by a suffix. Spelling [got] with a ⟨д⟩ helps the reader associate it with other forms of the same lexeme and distinguish it from homophonous word forms with underlying [t], in this case the word for 'Goth', which retains the [t] before suffixes (11b).

- (11) a. год ['got] 'year', годы ['godɪ] 'years'
 b. гот ['got] 'Goth', готы ['gotɪ] 'Goths'

Russian has five vowel phonemes, each with two different spellings (Table 3.13), although the two dots on ⟨ё⟩ are usually omitted, making ⟨е⟩ ambiguous. There are two main reasons for having two spellings per vowel. First, the spellings on the right side of each pair in the table can stand for the sequence [j] plus vowel (12a). The letters on the left are never used that way. Second, the spellings on the right can indicate that the preceding consonant is palatalized (12b). But putting vowel symbols to multiple purposes can lead to conflict. How to mark palatalization when a palatalized consonant is not followed by a vowel? This problem is addressed by using ⟨ь⟩ as a silent mark of palatalization (12c). The same solution is applied to show that a [j] is sounded between a palatalized consonant and a vowel (12d): note that without the ⟨ь⟩, *ню would spell *['pʲu]. To spell a [j] between a nonpalatalized consonant and a vowel, the subtly different symbol ⟨ъ⟩ is used (12e), though an in-progress sound change palatalizing consonants before [j] is gradually eroding the need for this rule. Centuries ago, the symbols ⟨ь⟩ and ⟨ъ⟩ represented ultrashort vowels called ЈЕРС – [i̯] and [ɨ̯], respectively – but those sounds are no longer in the phonological inventory, leaving the symbols available for repurposing.

TABLE 3.13
**Vowel phonemes
 of Russian**

i ⟨ы, и⟩	u ⟨у, ю⟩
e ⟨э, е⟩	o ⟨о, ё⟩
a ⟨а, я⟩	

- (12) a. я [ja] 'I'
 b. гуляю [gu'lʲaju] 'stroll'
 c. дать ['datʲ] 'to give'
 d. пью [pʲju] 'I drink'
 e. съезд ['sʲest] 'congress'

Different rules apply for the Cyrillic letters ц, ш, ж, ч, and щ. They retain their basic sharpness or flatness regardless of what vowel letter follows them.

Stress in Russian is phonemic: in principle, the position of the stress has to be memorized for each polysyllabic word. Although there are rules that can help one predict the stress of words, they are very complicated and have many exceptions. Indeed, there are some minimal pairs for stress such as [mu'ka] 'flour' versus ['muka] 'torment'. Despite the fact that stress is phonemic, the location of the stress is not marked in Russian writing.

The five-way vowel contrast shown in Table 3.13 obtains only in stressed syllables. In unstressed syllables, the vowel system is reduced to a three-way contrast [a]–[i]–[u]. Ignoring some exceptions, underlying [o] merges with [a], which may further reduce to [i] after a palatalized consonant. Underlying [e] merges with [i]. Russian spelling handles this situation much as it does obstruent voicing assimilation, by representing underlying forms rather than the reduced pronunciation. Often these spellings give the reader a hint as to what other word forms a particular word may be related to, because different syllables are often stressed in different forms or derivatives of a word. For example, spelling the first vowel of [sʲi'mʲja] with an ⟨e⟩ instead of an ⟨и⟩ shows its connection with the plural form, семья (13a). Similarly, spelling the second vowel of the adverb [xara'ʃo] with an ⟨o⟩ instead of an ⟨а⟩ shows its connection with the adjective (13b). However, the first ⟨o⟩ of хорошо and хороший is never pronounced as [o] in this or any related word, and so its use instead of ⟨а⟩ is unlikely to help Russian readers at all.

- (13) a. семья [sʲi'mʲja] 'family'; семьи ['sʲemʲji] 'families'
 b. хорошо [xara'ʃo] 'well'; хороший [xa'roʃij] 'good'

We won't usually devote much space to the phonetic description of allophones in a language, but a few allophonies are particularly salient in Russian. In addition to merging with other phonemes, unstressed vowels in most situations are centralized, with [a], [i], and [u] being pronounced as [ə], [ɪ], and [ʊ], respectively. Vowels tend to be pronounced a bit more high or front next to sharp consonants. Conversely, velar consonants tend to become palatal consonants before the front vowels [i] and [e]. The phoneme [i] is retracted to the central vowel [ɪ] when it appears after a flat consonant. Flat consonants tend to be somewhat VELARIZED in Russian: pronounced with the back of the tongue

raised toward the soft palate. This velarization is particularly noticeable before [i], where the transition between the consonant and vowel may sound like a velar GLIDE [ɥ]. A glide, or semivowel, is a consonant that is phonetically similar to or identical to a vowel but is found in the onset or coda of a syllable instead of in the nucleus.

3.5.4 MORPHOLOGY

3.5.4.1 Nouns

Like English, Russian nouns take different suffixes depending on whether their number is singular or plural.

GENDER is a system of putting nouns in a small number of categories so that other words can succinctly indicate which noun they are syntactically associated with. English has a trace of a gender system in its system of personal pronouns: to refer back to nouns designating females, a form of the feminine pronoun *she* must be used; the masculine pronoun *he* refers to males, and the neuter pronoun *it* refers to objects not characterized by sex. That is, English pronouns AGREE with their antecedents with respect to their semantic gender.

Russian has the same three genders as English, but gender plays a much larger role in the grammar. For one thing, gender is basically a property of a noun, not of the object the noun signifies. To be sure, with rare exceptions, nouns naming males are masculine, nouns naming females are feminine, and many nouns naming INANIMATE (neither human nor animal) objects are neuter. But individual nouns can contravene that expectation, and those naming inanimate objects are, more often than not, masculine or feminine rather than neuter. The Russian word for ‘desk’, [stoʃ], is masculine, and the word for ‘book’, [knʲiɡa], is feminine. Both Russian and English inherited very similar word-based gender systems from Proto-Indo-European. In Old English, words had grammatical gender, but English transitioned to a meaning-based system over 600 years ago; Russian did not. Another important difference from English is that in Russian, not just pronouns but also adjectives must agree in gender with the noun they modify. They take different suffixes depending on which gender the modified noun has – that is, they are said to INFLECT, or change their form, for gender.

A Russian noun takes different forms to identify the type of grammatical relation it has with its HEAD – that is, the main word in the phrase, on which it is dependent. Such a system is called an *inflectional CASE* system; a noun is said to inflect, or change its form, to express case. English has only vestiges of a case system, which, like gender, is best preserved in its personal pronouns. The following table arranges the pronouns by person, number, and semantic gender. *Person* refers to which of the two participants in a conversation are included: first (1) if the speaker is included, else SECOND (2) if the addressee is included, else third (3). Note that the three genders have one-letter abbreviations: masculine (M), feminine (F), neuter (N).

(14)	Category	NOM	ACC	GEN
	1SG	I	me	my
	2	you	you	your
	3SG.M	he	him	his
	3SG.F	she	her	her
	3SG.N	it	it	its
	1PL	we	us	our
	3PL	they	them	their

If a pronoun is the subject of a verb, it inflects to show the nominative case: *I grill hamburgers*; *She jogs*; and so forth. If a pronoun is the object of a verb, it inflects to show accusative case: *A mosquito bit me*; *A dog chased her*. If a pronoun is dependent on another noun – a typical relationship is to show possession – then it is in the genitive case: *My dog has fleas*; *Mary admires her perseverance*.

In addition to illustrating that case is connected to grammatical relations, pondering the English pronouns also reveals some pervasive ambiguities when it comes to case. First, cases are very commonly used for more than one grammatical relation. Note that the accusative case may express not only the object of a verb but also the object of a preposition: *An apple fell on me*. Second, the same word form may express more than one case: note *you*, *her*, and *it*. Third, different words may take different suffixes to express the same case. Note how four of the pronouns have a final *r* in the genitive, but two of them have *s*.

Russian has six main cases, which apply to nouns as well as to pronouns. The most fundamental uses of the cases are illustrated in (15). For each case, we have given an example sentence putting the Russian word for ‘book’ in that case. You will readily notice that Russian nouns mostly have different forms in the different cases. In contrast, in only one case does an English noun have a distinctive form.

- (15)
- a. NOM = nominative: subject of the verb
[**knʲiga** u'paʃa v 'vodu] ‘The **book** fell in the water.’
 - b. ACC = accusative: direct object of the verb
[**on** 'vzʲaʃ **knʲigu**] ‘He took the **book**.’
 - c. DAT = dative: indirect object of the verb
[**on** 'daʃ **knʲigi** 'dva kan'tsa] ‘He gave the **book** two endings.’
 - d. GEN = genitive: dependent of another noun
[**abʲto**ʃka **knʲigi** bi'ʃa a'torvana] ‘The **book**’s cover has been torn off.’
 - e. INS = INSTRUMENTAL CASE: the instrument with which an action is performed
[**on** 'mn'e gra'zʲit **knʲigaj**] ‘He threatens me with a **book**.’
 - f. LOC = locative: object of preposition showing location
[**ja** pra'tʃi'taʃ 'eta f **knʲigi**] ‘I read this in a **book**.’

Each case has many other uses, of which the most common is that different prepositions select for different cases.

Russian is very representative of Indo-European languages in expressing case, number, and gender by means of suffixes. Suffixes that express grammatical categories like these are often simply called ENDINGS. Russian is a fusional language, in that individual endings inseparably express number, gender, and case. For example, the ending [a] in ['knʲig-a] 'book' cannot be broken into three parts each of which expresses one specific category. Rather, [a] itself inseparably marks nominative case, singular number, and feminine gender. Because inflectional endings are so short, it may come as no surprise that there is a lot of homophony. [a] can also mean nominative, plural, and neuter, as in ['bʲud-a] 'dishes'. Russian is also typical of fusional languages in that different nouns may take different endings to express the same categories. For instance, the instrumental singular of ['knʲig-a] 'book' is ['knʲig-aj], but for ['kosʲtʲ] 'bone' it is ['kosʲtʲ-ju]. Fortunately, most words can be grouped into one of a small number of inflectional PARADIGMS: sets of words that take the same endings. Table 3.14 shows the full set of inflections for the three most important noun paradigms. Other paradigms are mostly variants of one of these.

The entries in Table 3.14 are the forms that appear when the first vowel of the ending is stressed. For example, [t͡ɕir'ta] 'line' is stem [t͡ɕirt-] plus the nominative singular stressed ending [a]; its accusative singular is [t͡ɕir'tu], and its dative singular is [t͡ɕir'tʲe]. The palatalization on the stem-final consonant of this last word form is due to the ending: the notation [-ʲe] indicates that the ending causes

TABLE 3.14
Noun paradigms

	Paradigm		
	1	2	3
	SG		
NOM	-a	M -∅, N -o	-∅
ACC	-u	= NOM OF GEN ^a	-∅
DAT	-ʲe	-u	-i
GEN	-i	-a	-i
INS	-oj	-om	-ju
LOC	-ʲe	-e	-i
	PL		
NOM	-i	M -i, N -a	-i
ACC	= NOM OF GEN ^a	= NOM OF GEN ^a	= NOM OF GEN ^a
DAT	-am	-am	-am
GEN	-∅	M -of, N -∅	-ej
INS	-amʲi	-amʲi	-amʲi
LOC	-ax	-ax	-ax

^aIn animate nouns, form is same as genitive; in inanimate ones, same as nominative.

[ʲ]-palatalization on any consonant that has a [ʲ] counterpart. Not all nouns are stressed on the ending. Some are stressed on the stem; some are stressed on the stem in some categories and on the ending in others (recall 13a). When the endings are unstressed, the vowel reductions discussed in §3.5.3 may occur. Compare the paradigm 1 singular case endings in Table 3.14 with the example sentences given in (15). That shows the inflection forms taken by a paradigm 1 noun that does not stress the endings. The vowel [o] becomes [a] and [e] becomes [i]. After sharp consonants, many speakers further reduce [o] and [a] to [i], but not when [a] is the last vowel in the word.

The notation -∅ in the table means that the stem does not take any ending. When that happens, words ending in consonant clusters often insert an [e] or an [o] to break up the cluster. For example, the word [dʲnʲ-i] ‘days’, a nominative plural paradigm 2 form, would according to Table 3.14 have the nominative singular *[dʲnʲ]. Instead it inserts [e], resulting in [dʲenʲ]. Historically, all forms of the word had a jer. Jers were deleted in certain environments, and in others they became full vowels.

The notation “= NOM or GEN” in the accusative rows of Table 3.14 means that different nouns in the same declension take different endings, depending on whether the object they refer to is animate. People and animals take one set of endings; inanimate objects take another. In Russian, the category of animacy is determined by the meaning of the noun, and is not an arbitrary property of words, at least not nearly to the same extent that gender is.

3.5.4.2 Adjectives

Like nouns, adjectives have endings that indicate gender, case, and number (Table 3.15). An adjective must agree with the noun it modifies in those three categories, as well as animacy in the accusative case. Note that the adjective endings are mostly different from those of nouns; adjectives must match the nouns in terms of the function of their inflections, not their form.

The endings listed are for when they are stressed; two-syllable endings are stressed on their first vowel. For words in which the stem is stressed, the unstressed

TABLE 3.15

Sample adjective declension

Case	SG			
	F	M	N	PL
NOM	-aja	-oj	-oja	-iji
ACC	-uju	= NOM or GEN	-oja	= NOM or GEN
DAT	-oj	-omu	-omu	-im
GEN	-oj	-ova	-ova	-ix
INS	-oj	-im	-im	-imʲi
LOC	-oj	-om	-om	-ix

suffixes undergo the expected change of [o] into [a]. Exceptionally, the masculine nominative singular is nowadays mostly pronounced [ij] when unstressed.

Here are some examples of how adjectives inflect to agree with nouns. Sentence (16e) illustrates a special feature of adjectives in many Slavic languages. In predicate nominative constructions, adjectives often have forms with noun inflections, which are used as the head of nominative predicates. These forms – the singular has feminine [-a] (noun paradigm 1), masculine with no ending, and neuter singular in [-o] (both paradigm 2), and the plural has [-i] – are shorter than the more usual adjective endings, and so are commonly called *short* forms.

- (16) a. balʲf-oj 'konʲ
 big-M.NOM.SG horse(M)[NOM.SG]
 ‘a big horse’
- b. balʲf-aja 'koʃk-a
 big-F.NOM.SG cat(F)-F.NOM.SG
 ‘a big cat’
- c. balʲf-oja 'dʲerʲiv-a
 big-N.NOM.SG tree(N)-N.NOM.SG
 ‘a big tree’
- d. balʲf-im 'koʃk-am
 big-DAT.PL cat(F)-DAT.PL
 ‘to big cats’
- e. 'koʃk-a kra'sʲiv-a
 cat(F)-NOM.SG beautiful-F.NOM.SG
 ‘The cat is beautiful.’

3.5.4.3 Diminutives

Both nouns and adjectives in Russian have a variety of diminutive suffixes, which can sometimes be piled up one after another:

- (17) a. 'gorat
 city
 ‘town, city’
- b. gara'd-ok
 city-DIM
 ‘small town’

- c. gara¹d-ot̩̥-ik
 city-DIM-DIM
 ‘a tiny town’

[-ot̩̥] is an allomorph of [-ok], originally caused by palatalization. [-ik] is a different, though etymologically related, diminutive suffix.

- d. gara¹d-ot̩̥-it̩̥-ik
 city-DIM-DIM-DIM
 ‘a teensy-weensy town’

Besides indicating the small size of something, the diminutive also expresses endearment, and therefore plays a large role in baby talk and in the sweet talk between lovers. More rarely these days, it can also indicate condescension.

- (18) a. ¹p¹otr
 Peter(M)[M.NOM.SG]
 ‘Peter’

Underlying [e] often changes to [o] before flat consonants.

- b. ¹p¹et̩̥-in¹k-a
 Peter(M)-DIM-NOM.SG
 ‘my dear Petey’

Noun paradigm 1 (note the [-a]) is most commonly used for feminine words, but it includes quite a few masculine words as well. These are limited to words that refer to males (semantically masculine words).

- c. ¹p¹et̩̥-k-a
 Peter(M)-DIM-NOM.SG
 ‘Pete’

a servant, or someone of similarly low social status, or someone with whom the speaker is angry, annoyed, etc.

Adjectives can be given diminutive suffixes. The diminutive meaning applies to the modified noun, not to the adjective.

- (19) a. ¹b¹et̩̥-ij
 white-M.NOM.SG
 ‘white’

- b. ¹b¹et̩̥-in¹k-ij
 white-DIM-M.NOM.SG
 ‘tiny and white’ (not *‘whitish’)

3.5.4.4 Patronymics

A source of endless confusion for the readers of Russian novels in translation are **PATRONYMICS**, which Russians use as middle names. The patronymics name a person's father. It is considered respectful enough to address someone only by his or her first name followed by the patronymic, omitting the family name or any title of respect.

Patronymics as well as Russian family names are adjectival in form and must agree in gender with the nouns they modify, that is, with the sex of the bearer. This creates confusion among uninitiated non-Slavic readers because both the patronymic and the family name will look different in the case of a brother and sister:

- (20) a. i'van i'van-av^ji't̪ɕ̚
 Ivan(M)[M.NOM.SG] Ivan-patronym(M)[M.NOM.SG] Ivan¹n-of
 'Ivan Ivanov, son of Ivan'
- b. ji'l'ʲen-a i'van-avn-a iva¹n-ov-a
 Elen(F)-F.NOM.SG Ivan-patronym(F)-F.NOM.SG Ivan-surname-F.NOM.SG
 'Elena Ivanova, daughter of Ivan'

3.5.4.5 Verbs

FINITE FORMS

A **FINITE VERB** is the most fundamental type of verb inflection in a language. It can serve as the head of a sentence. In synthetic languages, it is the type of verb that tends to have the most grammatical information attached to it, such as tense (time) and agreement with the subject. For example, in the English sentence *John loves to play WoW*, *loves* is the finite verb: the head of the sentence, marked for present tense and for having a third-person singular subject. The verb *play*, by contrast, is **NONFINITE**.

Like most other Indo-European languages, Russian has finite verb forms that agree with the number and person of their subject nouns. The verbs mark agreement by means of inflectional endings. Table 3.16 gives two typical inflectional

TABLE 3.16

**Personal verb forms
in Russian**

Subject	Paradigm	
	1	2
1SG	-u	-u
2SG	-iʃ	-iʃ
3SG	-it	-it
1PL	-om	-im
2PL	-ite	-ite
3PL	-ut	-at

ASPECT

Russian has an unusually thoroughgoing treatment of ASPECT at the lexical level. Aspect characterizes the action of a verb in terms of its completion. A PERFECTIVE ASPECT (PFV) frames the action as complete (Latin *perfectum*), whereas an IMPERFECTIVE ASPECT (IPFV) frames the action or state as ongoing, or not completed. In English, sentences such as *I have written a letter*, *I had written a letter*, *I sneezed*, and *I will write a letter tomorrow* suggest a perfective aspect. Sentences such as *I am writing a letter*, *I was writing a letter*, *I used to write letters*, *I sneezed over and over*, *I have been writing letters*, and *I'll be writing a letter when the movie starts* suggest an imperfective aspect. Most Russian verbs have a specific aspect built into the lexeme: a verb is either perfective or imperfective and cannot be used to express the other aspect.

A hallmark of perfective verbs in Russian is that they cannot express present tense: the present is considered inherently imperfective. Instead, when a perfective verb is used in the nonpast personal form (as in Table 3.16), the tense it expresses is future. For example, the imperfective verb [p'i'su] in the nonpast tense is interpreted as present 'I write', but the perfective verb [nap'i'su] in the same inflected form is future: 'I will write'.

Russian makes extensive use of derivation to make fine distinctions of meaning, including differences in aspect. Almost all imperfective verbs have at least one perfective counterpart that is derived from the imperfective verb, or vice versa. For example, from the basic root [stuk-] 'knock' one can form the following verbs, which are all cited in their infinitive form:

- (22) a. [stu'tɕ-atʲ] IPFV 'to knock, keep on knocking'
 b. ['stuk-nutʲ] PFV 'to knock (once)'
 c. [pa-'stuk-ivatʲ] IPFV 'to keep on knocking intermittently'
 d. [pa-stu'tɕ-atʲ] PFV 'to knock intermittently once'
 e. [da-stu'tɕ-at-sa] PFV 'to knock until there is a result'
 f. [na-stu'tɕ-at-sa] PFV 'to have one's fill of knocking'
 g. [ras-stutɕ-atʲ] PFV 'to knock something apart'
 h. [ras-stutɕ-at-sa] PFV 'to knock away with abandon'

3.5.5 SYNTAX

The unmarked or basic word order in Russian is SVO, as in English. Adjectives normally precede the nouns they modify. However, the Russian word order is less fixed than that of English. The endings on nouns and their dependents indicate fairly unambiguously the grammatical relations: what modifies what, what is the subject of a verb, or the object, and so forth. The normal word order is often violated for stylistic reasons and when elements of the sentence are TOPICALIZED – marked as being the topic of discourse. For example, both of the following Russian

sentences basically mean ‘My cat bit your dog’, in spite of the radical difference in word order:

- (23) a. ma'j-a 'koʃk-a uku's'i-ɫ-a tva'j-u sa'bak-u
 my-F.NOM.SG cat-F.NOM.SG bite-PST-F.SG your-F.ACC.SG dog-F.ACC.SG
 Normal word order, perhaps in answer to the question “What happened?”
- b. tva'j-u sa'bak-u uku's'i-ɫ-a ma'j-a 'koʃk-a
 your-F.ACC.SG dog-F.ACC.SG bite-PST-F.SG my-F.NOM.SG cat-F.NOM.SG
 ‘Your dog’ placed at the front of the sentence, perhaps in answer to the question
 “What happened to my dog?”

Note that the Russian sentence meaning ‘Your dog bit my cat’ could have the same word order as (23b), but the nouns would have different endings:

- (24) tva'j-a sa'bak-a uku's'i-ɫ-a ma'j-u 'koʃk-u
 your-F.NOM.SG dog-F.NOM.SG bite-PST-F.SG my-F.ACC.SG cat-F.ACC.SG

The Russian equivalent of the verb ‘be’ is not used in the present tense, resulting in verbless clauses. Predicate adjectives used in such constructions may be used in their short form, if they have one.

- (25) a. 'ja stu'dʲent
 I[nom] student[M.NOM]
 ‘I am a student.’
- b. 'ja 'bolʲin
 I[nom] sick[M.NOM]
 ‘I am sick.’
 ['bolʲin] is the short form of [balʲnoj].

3.5.5.1 Interrogative sentences

Interrogative sentences (questions) in Russian often have the same word order as DECLARATIVE SENTENCES (statements), especially in CONTENT QUESTIONS – those that have an interrogative word like ‘when’ that the answerer is meant to replace with the correct content:

- (26) a. 'ti 'bi-ɫ 'tam
 you be-PST[M.SG] there
 ‘You were there.’

- b. ka'gda 'ti 'bi-ʔ 'tam
 when you be-PST[M.SG] there
 'When were you there?'

Note how the subject and verb switch places in the English equivalent but not in the Russian.

In POLAR QUESTIONS (also called *yes-no questions*), the word that is the focus of the interrogation is fronted to the beginning of the sentence, and the enclitic interrogative PARTICLE [ʔi] is placed immediately after it. A particle (Q) is a word that conveys grammatical information and doesn't belong to any other major word class.

- (27) a. 'bi-ʔ=ʔi 'ti 'tam
 be-PST[M.SG]=Q you there
 'Were you (actually) there?'
- b. 'ti=ʔi 'bi-ʔ 'tam
 you=Q be-PST[M.SG] there
 'Was it you (and not somebody else) who were there?'

The use of the enclitic [ʔi] is now considered literary. In everyday language, the most common way of forming polar questions is to use the same words and the same word order as the declarative statement, but apply an interrogative intonation.

3.5.5.2 Subordinate clauses

A SUBORDINATE CLAUSE is a clause that is syntactically dependent on some element in a higher, parent clause. In (28), [ka'gda 'ti pr'i'ʃoʔ] 'when you came' is a subordinate clause that is dependent on and elaborates the tense of the verb ['biʔ] in the main clause ['on u'ʒe 'biʔ 'p'ʃan] 'he was already drunk'.

- (28) ka'gda 'ti pr'i'ʃoʔ 'on u'ʒe 'bi-ʔ 'p'ʃan
 when you come(PFV)-PST[M.SG] he already be-PST[M.SG] drunk[M.NOM.SG]
 'When you came, he was already drunk.'

A RELATIVE CLAUSE is a subordinate clause that is dependent on a noun. Relative clauses in Russian follow the noun they modify (29). Like English, Russian employs RELATIVE PRONOUNS (REL) as markers of relative clauses; they agree in number and gender with the noun they depend on. Their case, however, is determined by their usage within the relative clause itself.

- (29) 'koʃk-a ka'tor-aja uku's'i-ʔ-a ka'n'i-a
 cat-F.NOM.SG which(REL)-F.NOM.SG bite(PFV)-PST-F.SG horse-M.ACC.SG
 ub'i'ʒa-ʔ-a
 run_away(PFV)-PST-F.SG
 'The cat that bit the horse ran away.'

3.5.5.3 Negation

Like French, Spanish, and some other languages, including many nonstandard dialects of English, Russian employs double negation. The multiple negatives do not cancel each other out, but redundantly signal and reinforce the negative aspect of the statement:

- (30) 'on n'ika'gda n'i 'xotʃ-it paj-'ʔi n'iku'da sa 'mnoj
 he never not want-3SG go(PFV)-INF nowhere with me
 'He never wants to go anywhere with me.'

Another feature of negation in Russian is the different case marking of direct objects of negated verbs. The genitive case is used instead of the accusative case when the object is indefinite.

- (31) a. 'ja 'v'iʒ-u 'koʃk-u
 I see(IPFV)-NPST.1SG cat-F.ACC.SG
 'I see a cat.'
- b. ja n'i 'v'iʒ-u 'koʃk-i
 I not see(IPFV)-NPST.1SG cat-F.GEN.SG
 'I don't see any cat.'

The genitive case is used instead of the expected nominative case to mark the logical subject in existential sentences that have been negated:

- (32) a. va'd-a 'jest'
 water-F.NOM.SG is
 'There is water.'
- b. va'd-i 'n'et
 water-F.GEN.SG isn't
 'There isn't any water.'

3.5.6 SAMPLE TEXT

The following passage is taken from Aleksandr Solzhenitsyn's novel *В круге первом* [f 'krugʲi p'ervam] (Solzhenitsyn 1968: 95). Solzhenitsyn (Солженицын [salʒi'nʲitsin]) was a Nobel Prize-winning author best known for his depiction of Soviet forced labor camps. In this selection, Abakumov [aba'kumaf], Minister of State Security under Stalin, is interrogating Bobynin [ba'binʲin], an engineer who works in a secret research institute staffed with prisoners.

In order to keep things uncluttered, some morpheme breaks have been ignored.

- (33) У Абакумова был тоже голосок с громовыми раскатами, и он умел им припугнуть.

u aba'kumav-a bi-ʔ 'toʒi gala's-ok
at Abakumov-M.GEN.SG be-PST[M.SG] also voice-DIM[M.NOM.SG]

z gram-a'vi-mʲi ra'skat-amʲi
with thunder-ADJ-M.INS.PL roll-M.INS.PL

i 'on u'mʲe-ʔ i-m prʲipu'gnu-tʲ
and 3SG.M.NOM know_how(IPFV)-PST[M.SG] 3SG.M-INS frighten(PFV)-INF

'Abakumov also had quite a little voice, like rolling thunder, and he knew how to intimidate people with it'

The most common way to express possession in Russian is by using an existential construction (32). The possessor is the object of the preposition [u] 'at', which takes the genitive case.

The preposition [s], which has the allomorph [z], takes the instrumental case when it means 'with'. Unlike with [im] later in this sentence, this isn't the core case function of means. It is very common for instrumentality and accompaniment to be linked in case systems.

- (34) Но сейчас он чувствовал, что кричать было бы бесполезно и несолидно.

'no sʲi'tʲas 'on tʲustava-ʔ ʃta kpʲi'tʲ-atʲ
but now 3SG.M.NOM feel(IPFV)-PST[M.SG] that(SUBORD) shout-INF

'bi-ʔ-a=bi bʲis-'pomaʒ:n-a i nʲi-sa'lʲidn-a
be-PST-N.SG=SBJV without-help-N.NOM.SG and non-solid-N.NOM.SG

'But for now he felt that it would be useless and not respectable to shout.'

ADJ = ADJECTIVE.

SUBORD = SUBORDINATOR, a conjunction that introduces subordinate clauses.

SBJV = SUBJUNCTIVE MOOD, making the clause hypothetical.

Neuter nominative singular forms of adjectives can often be translated as impersonal constructions in English. Another example: [ʲxoladn-a] 'It is cold'. The same adjective form is also commonly used as an adverb in Russian.

- (35) Он понял, что арестант этот трудный.

'on 'pon'i-ɫ ʃta ar'i'stant
 3SG.M.NOM understand(PFV)-PST[M.SG] that(SUBORD) prisoner(M)[M.NOM.SG]
 'etat 'trudn-ij
 this[M.NOM.SG] difficult-M.NOM.SG

'He understood that this prisoner was difficult.'

- (36) И только предупредил: — Слушайте, заключённый.

i 'toɫka pr'iɫupr'i'di-ɫ 'sluʃa-j-tʃi zakɫ'u tʃonnij
 and only warn(PFV)-PST[M.SG] listen(IPFV)-IMP-PL prisoner(M.NOM.SG)

'And so he merely warned: "Listen, prisoner.'

[ʃluʃaj-tʃi]. The ending is plural, but the command is addressed to one person. Use of the plural in direct address was adopted as a more polite or formal way of addressing a single individual, like French *vous*. In Soviet Russia it was the normal way of addressing any adult who was not a close friend or family member.

The mark "—" is a quotation dash, used to set off quoted speech or indicate change of speakers.

- (37) Если я с вами мягко, так вы не забываетесь

'jes'ɫi 'ja s 'vam'i m'axk-a 'tak 'vi n'i zabi'va-j-tʃi-s'
 if 1SG.NOM with 2PL.INS soft-ADV so 2PL.NOM not forget(IPFV)-IMP-PL-REFL

'"Just because I'm going easy on you, don't you forget yourself. ..."

ADV = ADVERB.

REFL = REFLEXIVE VOICE. In Russian, the reflexive is indicated by attaching [sʲ] (sometimes [s'a] or [sa]) after all other inflectional endings. In a true reflexive, such as *John shaves himself*, the subject of a verb is understood to be the same entity as its object. As in many other languages, constructions that are formally reflexive in Russian are not always literally so from a semantic perspective, but all reflexive verbs do have the common property of not taking a direct object.

[ʃa s 'vam'i m'axk-a]. It is not unusual in Russian for a verb to be omitted if it can be supplied from context.

- (38) — А если бы вы со мной грубо – я б с вами и разговаривать не стал, гражданин министр.

a 'jes'ɫi=bi 'vi sa 'mnoj 'grub-a ja=p s 'vam'i i
 but if=SBJV 2PL.NOM with 1SG.INS harsh-ADV 1SG.NOM=SBJV with 2PL.INS even

razga'var'iva-tʃi n'i 'sta-ɫ graʒda'n'in m'i'n'istr
 talk(IPFV)-INF not begin(PFV)-PST[M.SG] citizen[M.NOM.SG] minister[M.NOM.SG]

'"If you were dealing with me harshly, I wouldn't even begin to talk to you, citizen minister.'

Bobyenin, being a prisoner, cannot address Abakumov as [ta'variɕ; m'i'n'istr] 'comrade minister', which would have been the normal way for an ordinary Soviet citizen to address such an official. Prisoners and others deemed enemies of the state could not be considered comrades of Soviet citizens.

- (39) Кричите на своих полковников да генералов, у них слишком много в жизни есть, им слишком жалко этого всего.

kr'i'tɕ-i-tʲi na sva-'ix pał'kovn'ik-af da gin'i'rał-af u
shout(IPFV)-IMP-PL at REFL.POSS-M.ACC.PL colonel-M.ACC.PL and general-M.ACC.PL at

n'i-x 's'lj'ɨ'kam 'mnoga v 'ziz'n'i-i jes'tʲ i-m 's'lj'ɨ'kam 'załk-a
3PL-GEN too much in life-F.DAT.SG is 3PL-DAT too sorry-N.NOM.SG

'eta-va fs'ji-'vo
this-N.GEN.SG all-n.gen.sg

‘‘Shout at your colonels and generals! They have too much in life, they’d be too sorry to lose it all.’’

[n'ix]. The third-person pronoun adds an initial [n] when it follows a preposition. Compare [im].

['zalka] 'sorry' expresses the experiencer in the dative case and the stimulus, the thing one is sorry about, in the genitive case.

- (40) — Сколько нужно – и вас заставим.

'skol'ka 'nuʒna i v-as za'stav'i-m
as_much_as necessary[N.NOM.SG] and 2PL-ACC force(PFV)-NPST.1PL

‘‘We’ll do whatever we have to; and we’ll use force on you.’’

- (41) — Ошибаетесь, гражданин министр!

aʃi'baɨ-itʲi-sʲ graʒda'n'in m'i'n'istr
err(PFV)-2PL-REFL citizen[M.NOM.SG] minister[M.NOM.SG]

‘‘You are mistaken, citizen minister.’’

On the use of the reflexive, compare the French expression *vous vous trompez* ‘you are mistaken’.

- (42) — И сильные глаз-а Бобынина сверкнули ненавистью.

i 's'lj'ni-ji gla'z-a ba'bin'in-a sv'ir'knu-l'i-i
and strong-NOM.PL eye(M)-M.NOM.PL Bobynin-M.GEN.SG flash(PFV)-PST-PL

'n'ɨnav'istʲ-ju
hatred-F.INS.SG

‘And Bobynin’s piercing eyes flashed with hatred.’

Quite a few masculine nouns form their nominative plural in [a], which is always stressed.

- (43) — У меня ничего нет, вы понимаете – нет ничего!

u m'i'nʲ-a nʲitʲe-i'vo 'nʲet v-i pa'nʲima-itʲi 'nʲet nʲitʲe-i'vo
 at 1SG-GEN nothing-N.GEN.SG isn't 2PL-NOM understand-NPST.2PL isn't nothing-N.GEN.SG
 ‘‘I have nothing, you understand – *nothing!*’

- (44) Жену мою и ребёнка вы уже не достанете – их взяла бомба.

ʒi'n-u ma'j-u i fʲi'bʲonk-a v-i u'ʒe nʲi da'stanʲ-itʲi
 wife-F.ACC.SG my-F.ACC.SG and child-M.ACC.SG 2PL-NOM already not get(PFV)-NPST.2PL
 i-x vʒ'a-ʲ-a 'bomb-a
 3PL-ACC take(PFV)-PST-F.SG bomb-F.NOM.SG

‘‘You can't get my wife and child any longer: a bomb took them.’

- (45) Родители мои – уже умерли.

ra'dʲitʲilʲ-i ma-'i u'ʒe 'umʲir-ʲ-i
 parent-M.NOM.PL 1SG.POSS-M.NOM.PL already die(PFV)-PST-PL

‘‘My parents have already died.’

- (46) Имущества у меня всего на земле – носовой платок, а комбинезон и вот бельё под ним без пуговиц (он обнажил грудь и показал) – казённое.

i'muʃ:istv-a u mʲi'nʲ-a fʲʲi-'vo na ʒimʲʲ-e nas-a'v-oj
 possession-N.GEN.SG at 1SG-GEN all-N.GEN.SG on earth-F.DAT.SG nose-ADJ-M.NOM.SG

pʲa'tok a kambʲinʲʲon i 'vot bʲilʲʲo
 kerchief[M.NOM.SG] whereas overalls[M.NOM.SG] and here underwear[M.NOM.SG]

'pod nʲi-m 'bʲes 'pugavʲits 'on abna'ʒi-ʲ
 under 3SG-INS without buttons[F.GEN.PL] 3SG[NOM.SG] bare(PFV)-PST[M.SG]

'grutʲ i paka'za-ʲ ka'zʲonn-aja
 chest[F.ACC.SG] and show(PFV)-PST[M.SG] state_owned-N.NOM.SG

‘‘My only earthly possession is a handkerchief; the overalls and this underwear beneath it that doesn't have any buttons’’ — he bared his chest and showed him —
 ‘‘are state-issued.’

According to Soviet prison regulations of that time, prisoners were forbidden to wear clothing with buttons.

- (47) Свободу вы у меня давно отняли, а вернуть её не в ваших силах, ибо её нет у вас самого.

sva'bod-u v-i u mʲi'nʲ-a dav'no 'otʲnʲi-ʲ-i a
 freedom-F.ACC.SG 2PL-NOM at 1SG-GEN long_ago take_away(PFV)-PST-PL but

vʲir'nu-tʲ i'j-o nʲi v 'vaʃ-ix 'sʲiʲ-ax 'iba i'j-o
 return(PFV)-INF 3SG-F.ACC not in 2PL-LOC.PL power-F.LOC.PL for(CONJ) 3SG-F.GEN

nʲet u ʲv-as sam-aʲvo
 isn't at 2PL-GEN self-M.GEN.SG

“Freedom you took from me long ago, but it’s not in your power to return it, for you don’t have it yourself.”

CONJ = CONJUNCTION

- (48) Лет мне отроду сорок два, сроку вы мне отсыпали двадцать пять, на каторге я уже был, в номерах ходил, и в наручниках, и с собаками, и в бригаде усиленного режима – чем ещё можете вы мне угрозить?

ʲʲet ʲmnʲ-e ʲot-radu ʲsorak ʲdva
 years[M.GEN.PL] 1SG-DAT from-birth(ADV) forty[NOM] two[NOM]
 ʲsrok-u ʲv-i ʲmnʲ-e atʲsipa-ʲʲ-i ʲdvatsitʲ ʲpʲatʲ
 term-M.PTV.SG 2PL-NOM 1SG-DAT pour(PFV)-PST-PL twenty[acc] five[acc]
 na ʲkatarg-i ja uʲʒe bi-ʲʲ
 in convict_labor-F.LOC.SG 1SG[nom] already be-PST[M.SG]
 v namʲʲʲr-ax xaʲdʲʲi-ʲʲ i v na-ʲrutʲe-nʲʲik-ax
 in number-M.LOC.PL go(IPFV)-PST[M.SG] also in on-hand-NOUN-M.LOC.PL
 i s saʲbak-amʲʲ
 and with dog-F.INS.PL
 i v brʲʲʲgadi-i uʲʲʲʲinn-ava rʲʲʲʲim-a
 and in brigade-F.LOC.SG strenuous-M.GEN.SG regimen-M.GEN.SG
 ʲʲe-o ʲʲe-o ʲmoʒ-itʲʲ v-i mnʲ-e ugraʲʲzi-tʲʲ
 what-INS.SG still can(IPFV)-NPST.2PL 2PL-NOM 1SG-DAT threaten(PFV)-INF

“I am 42 years old, and you’ve given me a 25-year sentence; I’ve already been at convict labor, walked around wearing numbers, and in handcuffs, and with dogs, and in a hard labor brigade; what else can you threaten me with?”

PTV = PARTITIVE CASE. Russian uses this case for objects of verbs when they are MASS NOUNS – nouns that name substances that are not ordinarily treated as counted items. The idea is that the verb isn’t affecting all of the substance (here, time) but only part of it. The partitive can be thought of as a subclass of the genitive case, because for most nouns the same form is used for the genitive and the partitive.

[ʲʲet]. Some paradigm 2 masculine nouns lack a genitive plural ending. This is particularly common in the context of counts and measures.

‘Pour’ in the sense of ‘pour out a measure of grain’. Here this verb is being used as slang meaning ‘to mete out (a prison term)’.

- (49) чего ещё лишить? Инженерной работы?

ʲʲe-iʲvo ʲʲʲe-o ʲʲʲʲʲi-tʲʲ inʒʲʲnʲʲern-aj raʲʲbot-i
 what-GEN.SG still deprive(PFV)-INF engineering-F.GEN.SG work-F.GEN.SG

“What else can you deprive me of? Engineering work?”

- (50) Вы от этого потеряете больше.

'v-i 'ot 'eta-va pa'ti'rʲa-itʲi 'bolʲʲi
 2PL-NOM from this-N.GEN.SG lose(PFV)-NPST.2PL more

‘‘You would lose more from that.’’

- (51) Я закурю.

'ja zaku'rʲ-u
 1SG[nom] smoke(PFV)-NPST.1SG

‘‘I’m going to have a smoke.’’

- (52) Абакумов раскрыл коробку »Тройки« особого выпуска и пододвинул Бобынину:

aba'kumaf ras'kri-ʔ ka'ropk-u 'trojk-i
 Abakumov[M.NOM.SG] open(PFV)-PST[M.SG] box-F.ACC.SG Troika-F.GEN.SG
 a'sob-ava 'vipusk-a i pada'dvʲinu-ʔ ba'binʲin-u
 special-M.GEN.SG edition-M.GEN.SG and approach(PFV)-PST[M.SG] Bobyinin-M.DAT.SG

‘Abakumov opened a pack of Troika Special Edition, and pushed it over to Bobyinin.’

Higher-quality versions of cigarette brands and some other products were made for higher Soviet officials and were available only to them.

- (53) — Вот, возьмите этих.

'vot va'zʲmʲ-i-tʲi 'etʲ-ix
 here take(PFV)-IMP-PL this-F.PTV.PL

‘‘Here, take some of these.’’

This is perhaps a clearer use of the partitive object: ‘take (some) of these’, but not all of them.

- (54) — Спасибо. Не меняю марки. Кашель.

spa'sʲiba nʲi mʲi'nʲa-ju 'mark-i 'kaʲʲiʲ
 thanks not change(IPFV)-NPST.1SG brand-NOM.PL cough[M.NOM.SG]

‘‘Thanks. I’m not switching brands. The cough.’’

- (55) — И достал »беломорину« из самодельного портсигара.

i da'sta-ʔ bʲiʲa-'morʲ-in-u is sama-'dʲelʲ-n-ava
 and get(PFV)-PST[M.SG] white-sea-AUG-F.ACC.SG from self-make-ADJ-M.GEN.SG
 portsʲi'gar-a
 cigarette_case-M.GEN.SG

‘And he took out a coarse Belomor from a homemade cigarette case.’

AUG = augmentative word: the opposite of a diminutive, but much more often it is used to suggest coarseness, crudeness, and commonplaceness instead of large size. Bobyinin, of course, is baiting the official. Belomorkanal was one of the cheapest brands of cigarettes and had no filter; it defied common wisdom to suggest that it would be less likely to induce a cough than the minister’s fancy cigarettes.

- (56) — Вообще, поймите и передайте там, кому кому надо выше, что вы сильны лишь постольку, поскольку отбираете от людей не всё.

vaap'ɕe paj'm¹-i-t'i i p'ir'i'da-j-t'i 'tam ka-'mu
 in_general understand-IMP-PL and report(PFV)-IMP-PL there who(REL)-DAT.SG
 ka-'mu 'nada 'viʃi ʃta v-i s'il¹n-i
 who(REL)-DAT.SG necessary higher that(SUBORD) 1PL-NOM powerful-M.PL
 'l'iʃ pa'stol¹ku pa'skol¹ku ad-b'i'ra-it'i at l'u'd¹-ej
 only insomuch inasmuch from-take(IPFV)-NPST.2PL from people-M.GEN.PL
 n'i 'fs¹-o
 not all-N.ACC.SG

‘‘Understand this and report it to whichever higher-ups you need to, that you are powerful only as long as you don’t take *everything* away from people.’

- (57) Но человек, у которого вы отобрали всё – уже не подвластен вам, он снова свободен.

no tɕiʃa'v'ek u ka'tor-ava v-i at-a'bra-l¹-i
 but man[M.NOM.SG] at which-M.GEN.SG you-NOM.PL from-(PFV)-PST-PL
 'fs¹-o u'ʒe n'i pad-'vʃas¹t¹-in v-am
 all-N.ACC.SG already not under-power-ADJ[M.NOM.SG] 2PL-DAT
 'on 'sнова sva'bod¹-in
 3SG[M.NOM.SG] again free-ADJ[M.NOM.SG]

‘But a man you’ve taken *everything* from is no longer in your power; he is free again.’

3.6 Sketch of Finnish

3.6.1 GENERAL BACKGROUND

Finnish is a member of the Finnic clade of the Uralic language family. To a large extent, it is mutually intelligible with Karelian, which is spoken across the border in the Russian Republic of Karelia. Along with Swedish, Finnish is one of the two official languages of Finland, and it is spoken by some five million people.

Typologically, Finnish is predominantly agglutinative, with a very rich morphology. While most Uralic languages have SOV word order, Finnish has SVO.

Finnish vocabulary, too, has been influenced very much by Indo-European languages (58). Early Germanic loanwords in Finnish are important for the reconstruction of the Proto-Germanic language. The word [kuniŋ:as] (58b) shows the hypothesized Proto-Germanic form of the masculine nominative singular ending *[-az] more clearly than any attested word in a Germanic language.

- (58) a. [sata] ‘hundred’ < Iranian (cf. Avestan [satəm])
 b. [kuniŋ:as] ‘king’ < Germanic *[kuniŋgaz] (cf. Old English [kyning])
 c. [silta] ‘bridge’ < Lithuanian [tʃiʃtės]

- d. [risti] ‘cross’ < East Slavic [krʲist] (cf. modern Russian [krʲest])
- e. [joulu] ‘Christmas, Yule’ < North Germanic (cf. Swedish [ju:l])
- f. [ranta] ‘beach’ < Swedish [strand]
- g. [pihvi] ‘steak’ < English [bif] *beef*

The earliest written records in Finnish, dating back to the 1530s, are primarily Christian religious materials. Finnish has always used the Latin alphabet, including a few letters borrowed from Swedish: ⟨ä⟩ and ⟨ö⟩.

3.6.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

3.6.2.1 Phonemes

Table 3.17 lists the consonants that appear in native Finnish words; several additional sounds are found in relatively recent loanwords. The consonants are spelled as in IPA, except as indicated in angled quotes.

All of the native consonants except for [v], [j], [h], and [d] have phonologically long counterparts. Long consonants are written double in the orthography; for example, [t:] is ⟨tt⟩. This doubling of letters representing long consonants goes back to Greek and Latin spelling and is the source of the word GEMINATION, which comes from Latin *gemini* ‘twins’. Exceptionally, long [ŋ:] is written ⟨ng⟩, as in ⟨hangata⟩ [haŋ:ata] ‘to scrub’. A short [ŋ] phone appears only before [k], where [n] never occurs; as does the spelling ⟨n⟩, we treat short [ŋ] as an allophone of [n]. Finnish stops are not ASPIRATED – they are not followed by a noticeable period of voicelessness, as found after English voiceless stops. Voiced stops are not found initially in native words; [d] occurs only medially.

TABLE 3.17

Consonant phonemes of Finnish

Stops	p	t		k
· Voiced		d		
Fricative			s	
Nasal	m	n		ŋ: ⟨ng⟩
Trill			r	
Approximant	v ⟨v⟩		l j	h

TABLE 3.18

Vowel phonemes of Finnish

Height	Front		Back	
	Unround	Round	Unround	Round
High	i	y		u
Mid	e	ø ⟨ö⟩		o
Low	æ ⟨ä⟩		ɑ ⟨a⟩	

Finnish has a QUADRANGULAR VOWEL SYSTEM, as shown in Table 3.18: that is, even the low vowels have the contrast front versus back. There is a phonemic contrast between short and long for all vowels; the latter are written double in the orthography.

Finnish also has many diphthongs, which are written as a straightforward sequence of the two component vowel sounds: [ɑi] is ⟨ai⟩, [eu] is ⟨eu⟩, and so forth. In addition to many closing diphthongs, which end in a high vowel, Finnish also has three opening diphthongs, a type which is less common across languages: [ie], [yø], and [uo] as in the language's endonym, [suomi].

Finnish is justifiably cited as a language with an almost perfectly phonemic orthography: each letter unambiguously spells a different phoneme. There are, however, a few exceptions. The letter ⟨n⟩, normally [n], represents the bilabial [m] before ⟨p⟩ (⟨toisinpäin⟩ [toisimpæin] 'vice versa'). In addition, there are many situations where long consonants are written with only one letter, the same as the corresponding short consonant. This happens systematically when certain morphemes occur before a word that begins with a consonant. For example, *tule tänne* 'come here' is pronounced [tulet:æ̃n:e] (§3.6.3.4). Historically, words like *tule* ended in a consonant, but nowadays the only manifestation of that consonant is in how it affects other sounds. It is convenient to think of such words as ending in a silent consonant, because several other phonological processes also treat such words as if they end in consonants (e.g. gradation, §3.6.2.3).

Finnish is not a tone language. Stress is fixed, always on the first syllable of the word. Thus it is not lexically contrastive, and so will not be marked in our phonemic transcriptions.

3.6.2.2 Vowel harmony

Finnish words undergo front-back vowel harmony. As Table 3.18 shows, Finnish contains three front vowels that are identical to Finnish back vowels in height and rounding – all respects other than backness: [æ], [ø], and [y]. We call them the *harmonizing* front vowels, because they change from a front to a back articulation in order to harmonize, that is, agree, with other back vowels in the word. This assimilation occurs regardless of whether the vowels are short, long, or in a diphthong. Thus there are words like [poika] 'boy', but there couldn't be a word like *^χ[poiky], because that would put a back vowel ([o]) in the same word as one of the harmonizing front vowels. The front vowels that do not have a back counterpart, [i] and [e], do not harmonize, which is why [i] can exist in [poika].

Front-back vowel harmony must be obeyed even when new word forms are formed by adding suffixes. For example, the notion of being inside something is typically conveyed by adding the inflectional suffix [-s:æ̃] (see under "inessive case" in §3.6.3), as in [jæ̃rvis:æ̃] 'in a lake'. But trying to add the same suffix to [suome] 'Finland' would result in a harmony clash: the front vowel [æ̃] can't appear in the same word as the back vowels [u] and [o]. Therefore, the vowel of the suffix must be replaced with the corresponding back vowel: [suomes:ɑ̃]. Likewise, [ø] in a

suffix will become [o] if necessary, and [y] will become [u]. The rules do not apply between roots in compound words. Nor do they apply to recent loanwords, at least not in the formal standard language. For example, the loanword for ‘catalysis’ is [kataly:si], which mixes back [ɑ] with front [y]. See also the exercise on adaptation of loanwords in Finnish (§3.7.2).

3.6.2.3 Consonant gradation

Compare the nominative singular and genitive singular forms of the nouns in Table 3.19. The genitive singular is formed by adding the suffix [n] to the nominative singular. When [n] is added to a word that ends in a vowel, a formerly open syllable becomes a closed syllable: that is, a syllable that contains a coda consonant. That triggers a phonological change in Finnish: a stop undergoes weakening, or LENITION, when it is in the ONSET (prevocalic position) of a closed syllable. Lenition is quite common when a consonant occurs between vowels, but Finnish is unusual in restricting the process to the onset of closed syllables.

Rows 1 through 3 in Table 3.19 illustrate one common type of lenition: shortening of long consonants. Rows 4 and 5 illustrate intervocalic VOICING, another common lenition: [t] becomes [d]. It is important to note that Finnish doesn’t forbid [t] from occurring between voiced sounds in the onset of a closed syllable; [jutun] (row 2) is proof that a [t] is legal in that environment. Rather, whatever consonant appears in open syllables becomes one step more lenited in closed syllables. Because this is a stepwise, or graded, process in Finnish, it is called CONSONANT GRADATION. The change [t:] > [t] is one step, and [t] > [d] is another step. Importantly, the steps do not feed into each other in the derivation of the

TABLE 3.19
Consonant gradation in Finnish nouns

	Gloss	NOM	GEN
1.	‘soup’	sop:a	sop:a-n
2.	‘tale’	jut:u	jut:u-n
3.	‘cake’	kak:u	kak:u-n
4.	‘knowledge’	tieto	tiedo-n
5.	‘mother’	æiti	æidi-n
6.	‘bread’	leipæ	leivæ-n
7.	‘handle’	ripa	riv:a-n
8.	‘foot’	jalka	jala-n
9.	‘pig’	sika	sia-n
10.	‘chest’	rinta	rin:a-n
11.	‘town’	kaupunki	kaupun̩:i-n
12.	‘power’	valta	val:a-n
13.	‘occasion’	kerta	ker:a-n
14.	‘car’	auto	auto-n

same word. After *[jut:un] becomes [jutu-n], [jutu-n] doesn't go on to become *[judu-n].

You might expect that the other stops, [p] and [k], would become voiced stops as well, but in modern Standard Finnish, the consonant [p] actually lenites to the approximant [v] (rows 6 and 7). The consonant [k], in most cases, lenites to nothing at all (rows 8 and 9), the ultimate lenition: deletion.

Gradation can be complicated by other factors, such as the identity of the phonemes surrounding the consonant that undergoes lenition. Rows 10 through 13 illustrate one of the more important cases: the original stop may assimilate to the preceding consonant. The fact that that consonant becomes long is the only indication that the stop ever existed.

Gradation is not perfectly regular in Finnish, at least not on the surface. It does not apply in loanwords (row 14 in Table 3.19) and given names. Some words even seem to undergo reverse gradation: [tiede] 'science' becomes [tiete:n] in the genitive. The pattern becomes clear if we posit that [tiede] has a final silent consonant, which we will represent here as C_\emptyset . That is, the nominative is underlyingly [tiete C_\emptyset], with the word-medial stop being in a nominally closed syllable, thus triggering consonant gradation of [t] to [d], while the genitive is [tiete C_\emptyset en], with the word-medial stop being in an open syllable, which does not trigger gradation. Words with reverse gradation actually did formerly end in a consonant, and they retained their original gradation pattern after the consonant was lost.

3.6.3 MORPHOLOGY

Finnish words tend to be much longer than English words. Like many agglutinative languages, Finnish regularly strings several suffixes together at the ends of words. Also, root morphemes tend to be two syllables long, while roots of one syllable are favored in English.

3.6.3.1 Nouns

INFLECTION

Nouns in Finnish inflect for number, case, and possession, using separate morphemes in that order:

- (59) ravintolo-i-s:a-ni
 restaurant-PL-INE-1SG.POSS
 'in my restaurants'

The suffix glossed INE is an ending for the *INESSIVE CASE*, which is used for denoting location within something. The case system is discussed in more detail in the next paragraph. The plural marker is [-t] at the end of a word, but [-i-] word-medially and [-j-] between two vowels. Singular number is unmarked in Finnish and will normally be omitted from our glosses. This separation of case and

number markers is a different situation from Russian, a typical fusional language in which case and number are inseparably expressed by a single ending. The separability of number and case in Finnish is an important factor leading linguists to classify it as an agglutinative language.

CASES

Finnish is famous for having a very large number of cases, but not all are equally PRODUCTIVE. Nonproductive cases tend to be found only in fixed expressions. The most commonly used cases are listed in Table 3.20.

Nouns are subject to different phonological changes conditioned by different endings. You have already seen in Table 3.19 how the genitive singular stem can look quite different from that of the nominative singular because of consonant gradation; many other endings have the same effect. Gradation and other changes can take place in the other cases as well. Due in part to the fact that in the nominative singular the stem appears at the very end of a word, phonological processes may make it turn out quite different from the way the stem appears before endings. For example, the stem [rukoukse-] ‘prayer’ appears as [rukous] in the nominative singular. In addition, the endings themselves are subject not only to vowel harmony but also to other changes conditioned by the phonetic environment within a word; only a few of the variants produced by these changes are listed in Table 3.20. The complicated allomorphy within words makes Finnish appear more fusional than is typical of agglutinative languages.

The main functions of the nominative, accusative, and genitive cases are similar to those of Indo-European languages like Russian, marking grammatical relations between the parts of the sentence:

TABLE 3.20
Finnish cases

	Case	Suffix	Function
NOM	nominative	∅	subject
ACC	accusative	-n, ∅	object
PTV	partitive	-tæ, -æ	partial object
GEN	genitive	-n	modify noun
ESS	essive	-næ	‘as’, ‘at’
TRANSL	translative	-ksi	new state
INE	inessive	-s:tæ	‘in’
ILL	illative	-n, -sen, -hin	‘into’
ELA	elative	-stæ	‘out of’
ADE	adessive	-l:tæ	‘on’, ‘at’
ALL	allative	-l:e	‘onto’
ABL	ablative	-ltæ	‘from off of’
ABE	abessive	-t:tæ	‘without’
COM	comitative	-i-ne	‘with’

- (60) *pojā-n kis:a syø hirre-t*
 boy-GEN cat[nom] eat(3SG) mouse-ACC.PL
 POSSESSOR SUBJECT OBJECT
 ‘The boy’s cat will eat the mice.’

The partitive case has many uses, most of them conveying literally or metaphorically the idea of partialness. Very often the direct object of a verb is in the partitive case. The distinction between the accusative and the partitive is one of TELICITY. A telic action is one that expresses a complete activity that can be defined in terms of a specific result or goal. An object of a telic action is put in the accusative case. In (60), the clear implication is that the cat will eat up all the mice, then her project will be complete. As with many other telic sentences, there is an implication that the verb, which morphologically could be either present or future, should be interpreted as future, because it seems unlikely that the cat is going to finish her project right now. An ATELIC action is not directed toward a result, and its object is put in the partitive case. If (60) had put the word for ‘mice’ in the partitive – [hir-i-æ] – the implication would be that the cat is engaged in general mouse-eating activity, not with a goal of eating a specific set of mice, and the hearer would be more inclined to interpret the verb as a present-tense activity: ‘The boy’s cat is eating some mice.’ The partitive case is used to mark the direct object of negated verbs, verbs that indicate an ongoing action, or verbs that express thoughts, wishes, or hopes – all atelic activities or states.

The other Finnish cases fundamentally express location and are called *locative* cases. The basic meanings of these cases can be gathered from the English gloss in Table 3.20. Generally speaking, a locative case is functionally equivalent to a preposition in English.

- (61) a. *hæn on suome-s:a*
 3SG is Finland-INE
 ‘She (or he) is in Finland.’
- b. *tule-n kylæ-stæ*
 come-1SG village-ELA
 ‘I come from a village.’
- c. *kirjā on pøydæ-l:æ*
 book is table-ADE
 ‘There’s a book on the table.’
- d. *pāne-n kirjā-n pøydæ-l:e*
 put-1SG book-ACC table-ALL
 ‘I will put the book on the table.’

Even the partitive originally expressed a locative idea, ‘from.’ This meaning is still seen literally in the expression [koto-a] ‘from home’ and metaphorically in the atelic usages described in the preceding paragraph: if the cat eats *some* mice, it is eating a subset taken *from* all possible mice.

Six of the most commonly used cases express location according to two factors. The first factor is whether the reference point is inside or outside the referenced object. The INTERNAL CASES are the INESSIVE ‘in’, the ILLATIVE ‘into’, and the ELATIVE ‘(coming) from inside’ (i.e. ‘out of’). The EXTERNAL CASES are the ADRESSIVE ‘at’, the ALLATIVE ‘onto’, and the ABLATIVE ‘from’. The second factor is whether motion is involved and, if so, whether it is toward or away from the object. The inessive and adessive don’t indicate movement; the illative and allative show motion toward an object; and the elative and ablative show motion from an object. Originally these two factors were indicated by a sequence of two separate suffixes, which can today be partially discerned by noting the parallelism between four of the cases:

(62)	Movement	Internal	External
	From	ELA -stæ	ABL -ltæ
	None	INE -s:iæ	ADE -l:iæ

The compositionality of these four cases is reminiscent of the situation in Tabassaran (§3.3.2), whose formidably large number of cases turned out to be caused by pairing location suffixes with motion suffixes.

More often than not, locative cases express a variety of concepts that are only metaphorically or incidentally connected with literal position or movement in space. Here are a few examples of nonliteral use of Finnish locative cases:

- (63) a. poika-na minæ en tunte-nut hæn-tæ
 boy-ESS 1SG NEG.1SG know-PST.PTCP 3SG-PTV
 ‘As a boy I did not know him.’
- b. minu-l:a on kirja
 1SG-ADE be[3sg] book
 ‘I have a book.’

This latter sentence (63b) is very similar to the expression of possession in Russian (33).

The allative can be used to indicate an indirect object:

- (64) me hanki-m:e te-i-l:e u:de-t sukse-t
 1PL get-1PL 2PL-PL-ALL new-ACC.PL ski-ACC.PL
 ‘We will get you new skis.’

The ablative may mark loss, lack, or deficiency in general:

- (65) hæne-ltæ kuol-i æiti
 3SG-ABL die-3SG.PST mother
 ‘His mother died.’

The elative indicates the stimulus role of the verb ‘like’:

- (66) pidæ-t:e-kø hæne-stæ
 like-2PL-Q 3SG-ELA
 ‘Do you like her?’

If the multiplicity of functions for each case seems excessive – and we have given only a tiny sample of those that have been catalogued – it is useful to draw a parallel between Finnish locative cases and the English prepositions they correspond to. When someone says he believes *in* ghosts or is working hard *on* a term paper, the meanings of English prepositions are far from literal.

POSSESSIVE SUFFIXES

The last element that can be attached to nouns is a suffix that indicates the number and person of the possessor. For example, [talo-s:a-ni] ‘house-INE-POSS.1SG’ means ‘in my house’. There are separate forms for each combination of person and number, except that the third person does not distinguish number:

- (67) Person SG PL
 1 -ni -m:e
 2 -si -n:e
 3 -nsæ

The final consonant of any case or number ending is dropped before adding the personal suffix, which can lead to ambiguity. The words [talo] ‘house[NOM.SG]’, [talo-n] ‘house-GEN.SG’, and [talo-t] ‘house-NOM.PL’ all become [taloni] when [-ni] POSS.1SG is added. Possessive suffixes are an important part of the standard written language but aren’t used much at all in everyday speech.

Because the nominative case is unmarked in Finnish, we will henceforth normally omit it from our glosses: if case is not noted, it can be assumed to be nominative.

3.6.3.2 Adjectives

Adjectives precede nouns that they modify in the same noun phrase. As in Russian (§3.5.4.2), adjectives agree with the noun they modify in number and case. Unlike in Russian, they do not agree in gender, because there is no grammatical gender in Finnish. The endings, consonant gradations, and so forth, are the same as for nouns.

3.6.3.3 Personal pronouns

The personal pronouns are as follows; stem forms used before suffixes are in parentheses:

(68)	Person	SG	PL
	1	minæ (minu-)	me (me-i-)
	2	sinæ (sinu-)	te (te-i-)
	3	hæn (hæne-)	he (he-i-)

The plural pronouns add an additional suffix [-tæ] in the accusative and genitive cases, such as [me-i-dæ-n] ‘our’.

The first and second person pronoun subjects are usually omitted in the standard, formal language. This does not raise ambiguities, because Finnish verbs are marked to show the person and number of their subject. As noted for nouns, Finnish does not have grammatical gender. There aren’t even separate pronouns for distinguishing semantic gender, like English ‘he’ and ‘she’.

As in Russian (36), the second person plural can be used to express respect and formality when addressing a single person.

3.6.3.4 Verbs

FINITE FORMS

The Finnish verb paradigm has several finite forms. They are inflected to show the person and number of the subject, as are the nonpast verbs in Russian (§3.5.4.5).

The basic nonpast inflections are used to express both present and future tenses. Factors such as the choice between accusative and partitive cases for the object can help signal tense: the nonpast of telic verbs tends to be interpreted as a future (60). This is evocative of how the nonpast of perfective verbs are interpreted as future tense in Russian, as might be expected because telicity and perfectivity are closely related concepts. Table 3.21 shows the personal endings taken by verbs. When added directly to the verb stem, they indicate the nonpast tense: for example, [laula-n] means ‘I sing’ or ‘I will sing’. To express the past tense, an [i] is inserted before the personal ending: [laulo-i-n] < *[laula-i-n] ‘I sang, I was singing’. Because the nonpast is the unmarked tense, it will not normally be mentioned explicitly in our glosses.

The indicative (IND) mood has no special marker and will go unmentioned in our grammatical glosses. The CONDITIONAL MOOD (COND), which is used mostly

TABLE 3.21
**Personal endings
on Finnish verbs**

Person	SG	PL
1	-n	-m:e
2	-t	-t:e
3	-:	-væt

TABLE 3.22

Imperative mood in Finnish

Person	SG	PL
1	—	syø-kæ:-m:e 'let's eat!'
2	syøC _ø 'eat!'	syø-kæ: 'eat!'
3	syø-kø:-n 'let him eat!'	syø-kø:-t 'let them eat!'

to express hypotheticals, is made by suffixing [-isi] to the verb stem and then adding personal endings: [laula-isi-n] 'I would sing'. The POTENTIAL MOOD (POT) implies doubt or uncertainty or an assumption. It is made by suffixing [-ne] before the personal endings: [laula-ne-n] 'maybe I will sing'.

The imperative mood is formed quite differently (Table 3.22). The second person singular is the verb stem followed by a silent final consonant, which manifests itself by lengthening the initial consonant of the next word, if any. Neither the silent consonant nor the lengthening is indicated in the orthography. In the other forms there is an explicit imperative suffix, [-kæ:] or [-kø:], which is followed by special personal endings.

NONFINITE FORMS

Finnish forms four participles distinguished by tense and voice. These are built by adding suffixes to the stem, as in the following forms of the verb [syø-] 'eat':

(69)	ACT	PASS
	NPST syø- <i>v</i> æ 'eating'	syø- <i>t</i> æ- <i>v</i> æ 'being eaten'
	PST syø- <i>nyt</i> 'having eaten'	syø- <i>ty</i> 'eaten'

These can be used to modify nouns, in which case they inflect for case and number just like adjectives. They also are used to form analytic verb tenses. For example, PERFECT-TENSE versions of the tenses are formed by using [ole-] 'be' in the appropriate tense, as a finite auxiliary verb, followed by the past active participle: [olen laula-nut]. The perfect tense expresses past events that are still relevant in the present, much like the English perfect form *I have sung*.

Finnish is well known for having four or five different infinitive forms, but only one of them really corresponds in its functions to the infinitive form of English. The other infinitives are easily identified as GERUNDS, verbs that have some properties of nouns, because they take noun case endings after their characteristic suffix.

(70)	infinitive INF	-æ, -tæ	sano-a 'to say'
	gerund GER	-mæ-	+ ABE sano-ma-ta 'without saying'
		+ ILL	sano-ma:-n 'intending to say'
	time gerund	-e-	+ INE sano-e-s:a 'when saying'

NEGATIVE VERB

To express negation (NEG), Finnish uses a special auxiliary verb, which is inflected as any other Finnish verb. In the nonpast indicative:

(71)	Person	SG	PL
	1	e-n	e-m:ɛ
	2	e-t	e-t:ɛ
	3	ei	ei-væt

In the ordinary nonpast indicative, this negative auxiliary verb is combined with a special form of the verb, which is basically its stem followed by a final silent consonant, which causes consonant gradation:

(72)	e-n	tiedæ-C _∅
	NEG-1SG	know-NEG
		‘I don’t know.’

Other tenses and moods are expressed by modifying the content verb, not the negative auxiliary.

IMPERSONAL VERBS

Finnish finite verbs have a special IMPERSONAL form (IMPRS) that is used when one does not know or does not wish to express the subject of the verb. Impersonal verbs have a special suffix that expresses tense and mood, but they are not inflected for the person or number of the subject, because the sentence has no subject. Such sentences can be most literally translated into English by supplying the subject *one*, but it often sounds more natural to render the sentence as a passive in English.

(73)	kirje-i-tæ	sɑ:ti:n	sæ:n:øli:isesti
	letter-PL-PTV	receive-IMPRS.PST	regularly
			‘One received letters regularly.’ = ‘Letters were received regularly.’

Finnish does not have a construction equivalent to *Mary was kissed by John*, where the subject is expressed but demoted to the object of a preposition.

3.6.4 SYNTAX

3.6.4.1 Agreement

In Finnish, adjectives and DETERMINERS must agree in number and case with their head nouns. Determiners comprise a broad class of grammatical words that help to limit or identify the reference without modifying it: in English, this includes words like ‘that’, ‘the’, ‘our’, ‘some’, and ‘two’. Verbs must agree in person and number with

their subject. In the spoken language, however, number agreement is becoming increasingly restricted. Instead of using a first person plural verb with the pronoun [me] ‘we’, the spoken language uses an impersonal verb, which has no number inflection: [me puhutɑ:n] ‘we speak’, but literally, ‘we, one speaks’. And instead of using the distinctive third person plural of verbs, the spoken language extends the third person singular to be used with third person subjects regardless of number: [tytöt laula:] instead of [tytöt laulavat] ‘The girls are singing’.

3.6.4.2 Interrogative sentences

Finnish interrogative construction is very similar to that of Russian (§3.5.5.1). Content questions do not have any special marking.

- (74) kuka tule-:
 who come-3SG
 ‘Who is coming?’

In polar questions, [=kø] is added to the word that is being questioned, and that word is moved to the beginning of the sentence, just as [Pi] is used in literary Russian:

- (75) a. tule-:=ko hæn
 come-3SG=Q 3SG
 ‘Is she coming?’
 I.e. ‘Is she coming or not?’
- b. hæn=kø tule-:
 3SG=Q come-3SG
 ‘Is she coming?’
 I.e. ‘Is it *she* that is coming and not someone else?’

Negative questions add [=kø] to the verb of negation:

- (76) ei-væt=kø tiedæC_∅
 NEG-3PL=Q know-neg
 ‘Don’t they know?’

Subordinate polar clauses are marked with the interrogative particle, just like interrogative main clauses:

- (77) e-n tiedæ-C_∅ on=ko hæn suomalainen
 neg-1SG know-NEG be[3sg]=Q 3SG Finnish
 ‘I don’t know whether he is Finnish.’

3.6.4.3 Relative clauses

Relative clauses are introduced by relative pronouns, which cannot be omitted, as is sometimes done in English. The relative pronoun agrees with its antecedent in number but takes the case ending that is appropriate to its grammatical role in the subordinate clause:

- (78) kirje:-t jo-i-sta ker:o-i-n o-vat pøydæ:-læ
 letter-PL REL-PL-ELA talk-PST-1SG be-3PL table-ADE
 ‘The letters about which I was talking are on the table.’

3.6.5 SAMPLE TEXT

This passage, written by H. J. Viherjuuri, is taken from Austerlitz (1966: 42–43).

- (79) Meidän Suomalaisten Sauna
 me-idæn suoma-lais-t-en sauna
 1PL-GEN Finland-inhabitant-PL-GEN sauna
 ‘The Sauna of Us, the Finns’
- (80) Missä ikänä Suomessa kulkee ja näkee asunnon,
 mi:-s:æ ikænæ suome:-s:a kulke:- ja næke:- asun:o-n
 where-INE ever Finland-INE walk-3SG and see-3SG dwelling-ACC
 ‘No matter where in Finland one walks and sees a residence.’
 The third person singular form of the verb without an overtly expressed subject is often used in an impersonal sense: ‘one walks,’ ‘one sees.’
- (81) voi tarkkaava silmä havaita jonkin matkan päässä asuntorakennuksesta metsikön
 reunassa tai veden partaalla pienen hirsisen rakennuksen, saunan.
 voi tark:a:-va silmæ havai-ta jo-n-kin matka-n
 can[3sg] observe-ACT.PTCP eye perceive-INF what-GEN-some distance-GEN
 päe:-s:æ asunto-raken:ukse-sta metsi-kø-n reuna-s:a tai vede-n
 end-INE dwelling-building-ELA forest-DIM-GEN edge-INE or water-GEN
 parta:-l:a piene-n hirs-ise-n raken:ukse-n sauna-n
 shore-ADE small-ACC log-made_of-ACC building-ACC sauna-ACC
 ‘an observant eye may notice, at some distance from the residential building, at the edge of the woods or by the water’s edge, a small log building, the sauna.’
- (82) Sauna on suomalaisen kylpylaitos, ja jokaisella perheellä, köyhimmälläkin, on
 oma saunansa, jossa kylvetään vähintään kerran viikossa.
 sauna on suoma-laise-n kylpy-laitos ja jokaise-l:a
 sauna be[3sg] Finland-inhabitant-GEN bath-institution and every-ADE
 perhe:-l:æ køyh-im:æ:-l:æ-kin on oma sauna-nsa jo-s:a kylve-tæ:n
 family-ADE poor-SUPL-ADE-even be[3sg] own sauna-3SG REL-INE bathe-IMPRS

vähintäin ker:an vi:ko-s:a
at_least once week-INE

‘Sauna is the bathing institution of the Finns, and every family, even the poorest, has its own sauna where one bathes at least once a week.’

SUPL = SUPERLATIVE DEGREE of an adjective, like *-est* in English.

- (83) Vain kaupungeissa ja tiheissä asutuskeskuksissa sekä laitosten yhteydessä on yhteisiä saunoja.

vain kaupun;e-i-s:a ja tihe-i-s:æ asutus-keskuks-i-s:a sekæ
only city-PL-INE and dense-PL-INE settlement-center-PL-INE as_well_as
laitos-ten yhteyde-s:æ on yhteis-i-æ sauno-j-a
institution-GEN.PL connection-INE be[3sg] public-PL-PTV sauna-PL-PTV

‘Only in cities and densely populated areas and in connection with institutions are there public saunas.’

- (84) On vaikeata antaa niin täsmällistä määritelmää siitä, mikä suomalainen sauna on, että sen perimmäinen olemus sen avulla selviäisi,

on vaikea-ta anta-: ni:n täsmælis-tæ mæ:ritelmæ-: si-tæ
be[3sg] difficult-PTV give-INF such exact-PTV definition-PTV that-ELA
mikæ suoma-lainen sauna on et:æ se-n
which(REL) Finn-ADJ sauna be[3sg] that(SUBORD) it-GEN
perim:æinen ole-mus se-n avu-l:a selviæ-isi
ultimate be-NOUN it-GEN help-ADE clarify-COND[3sg]

‘It is difficult to give such a precise definition of what a Finnish sauna is that with its help one would clarify what its ultimate nature is,’

- (85) eikä se ainakaan suomalaisen omalta kannalta liene niin tarpeellistakaan,

ei-kä se ainaka:n suoma-laise-n oma-lta kan:a-lta
NEG[3sg]-and it anyway Finland-inhabitant-GEN OWN-ABL viewpoint-ABL
liene ni:n tarpel:is-ta-ka:n
be[POT.NEG] such necessary-PTV-even

‘and anyway, from a Finn’s own viewpoint at least such may not even be necessary,’

- (86) sillä suomalaiset ovat ainakin tuhannen vuoden ajan käyneet saunassa ryhtymättä sitä määrittelemään tai selittämään, ja silti se – ainakin sen vieläkin yleinen muoto savusauna – on yhä kautta vuosisatojen perusominaisuuksiltaan pysynyt samanlaisena.

sil:æ suoma-laise-t o-vat ainakin tuhan:e-n vuode-n
for(CONJ) Finland-inhabitant-PL be-3PL at_least thousand-GEN year-GEN
aja-n käy-ne:-t sauna-s:a ryhty-mæ-t:æ si-tæ
time-GEN visit-PST.ACT.PTCP-PL sauna-INE undertake-GER-ABE it-PTV
mæ:ritele-mæ-n tai selit:æ-mæ-n ja silti se ainakin se-n vieläkin
define-GER-ILL or explain-GER-ILL and yet it at_least it-GEN still

yleinen muoto savu-sauna on yhä kaut:a
 fundamental form smoke-sauna be[3sg] continually through

vuos-i-sato-j-en perus-ominaisu:ks-i-lta-n
 year-PL-hundred-PL-GEN basis-characteristic-PL-ABL-3SG

pysy-nyt sama-n-laise-na
 remain-PST.ACT.PTCP same-GEN-kind-ESS

‘for Finns have been visiting the sauna for at least a thousand years without undertaking to define or explain it, and yet it – at least its fundamental form, the smoke sauna – has remained the same throughout the centuries as far as its basic characteristics are concerned.’

[sama-n-laise-na] illustrates that some compounds contain a case marking between the two roots.

A smoke sauna is a sauna without a chimney, made of earth. The smoke must be cleared out before use.

- (87) ilman ulkoapäin tulleita kehoituksia ja opetuksia, ilman paljon maksavaa valistustyötä,
 ilman ulkoapäin tul-:e-i-ta kehoituks-i-a ja
 without from_outside come-PST.ACT.PTCP-PL-PTV advice-PL-PTV and
 opetuks-i-a ilman paljon maksa-ua-: valistus-työ-tä
 instruction-PL-PTV without much cost-PRS.ACT.PTCP-PTV enlightenment-work-PTV
 ‘Without advice or instruction that has come from outside, without much expensive enlightenment work,’

- (88) ilman apurahoja ja mallipiirustuksia
 ilman apu-raho-j-a ja mali-pi:rustuks-i-a
 without aid-fund-PL-PTV and pattern-drawing-PL-PTV
 ‘without grants-in-aid or blueprints,’

- (89) on jokainen kodin perustaja rakentanut itselleen saunan usein ennen kuin itse asumuksen.
 on jokainen kodi-n perusta-ja rakenta-nut itse-l:e-n
 be[3sg] every home-GEN found-AGT build-PST.ACT.PTCP self-ALL-3SG
 sauna-n usein en:en kuin itse asumukse-n
 sauna-GEN often earlier than self[acc] residence-ACC
 ‘every founder of a home has built for himself a sauna, very often before the residence itself.’
 AGT affixes turn a verb into a noun denoting a person who does that verb.

- (90) Saunan ympärille kiertyy suomalaisen elämä syntymästä kuolemaan asti:
 sauna-n ympäri-l:e kierty-: suoma-laise-n eläemä
 sauna-GEN around-ALL revolve-3SG Finland-inhabitant-GEN life

synty-mæ-stæ kuole-ma:n asti
 be_born-GER-ELA die-GER-ILL until

'A Finn's life revolves around the sauna from the time he is born to the time he dies.'

- (91) Siitä hän ammentaa terveyttä hyvinä päivinä, ja siitä hän hakee parannusta sairauksien sattuessa.

si:-tæ hæ n am:enta-: terveyt-:æ hyv-i-næ pæiv-i-næ ja si:-tæ hæ n
 it-ELA 3SG acquire-3SG health-PTV good-PL-ESS day-PL-ESS and it-ELA 3SG

hake-: paran:us-ta sairauks-i-en sat:u-e-s:a
 seek-3SG cure-PTV illness-PL-GEN happen-while-INE

'From it he acquires good health during good days and from it he seeks cures for his illnesses when they happen.'

- (92) Sauna merkitsee jokaiselle suomalaiselle nautintoa ja virkistystä.

sauna merkitse-: jokaise-l:e suoma-laise-l:e nautinto-a ja virkistys-tæ
 sauna mean-3SG every-ALL Finland-inhabitant-ALL enjoyment-PTV and recreation-PTV

'To every Finn the sauna means enjoyment and recreation.'

- (93) Melko varmasti näillä seikoilla on oma tärkeä osuutensa siihen, että sauna on elänyt ja jatkuvasti elää voimakkaana Suomen kansan elämässä.

melko varma-sti næ-i-l:æ seiko-i-l:a on oma tærkeæ
 somewhat clear-ADV this-PL-ADE factor-PL-ADE be[3sg] own important

osu:te-nsa si:-hen et:æ sauna on elæ-nyt ja
 role-3SG it-ILL that(SUBORD) sauna be[3sg] live-PST.ACT.PTCP and

jatkuva-sti elæ-: voimak:a-na suome-n kansa-n
 continuous-ADV live-3SG powerful-ESS Finland-GEN people-GEN

elæ-mæ-s:æ
 live-NOUN-INE

'It is fairly clear that each of these factors has its own important role, in that the sauna has been and continues to be a very powerful force in the life of the Finnish people.'

- (94) Sauna ei ole vain vältämätön terveydenlähde, vaan myös suuri ja korkeamman luokan nautintoväline.

sauna e-i ole vain væltæ-mæ-tøn terveyde-n-læhde va:n
 sauna NEG-3SG be[3sg]-NEG only dispense-GER-NEG health-GEN-source but

myös su:ri ja korkea-m:a-n luoka-n nautinto-væline
 also great and high-COMP-GEN type-GEN enjoyment-means

'Sauna is not only an indispensable source of health but is also a great and superior means of enjoyment.'

COMP = COMPARATIVE DEGREE of an adjective, which indicates a higher degree of the adjective than possessed by another entity. The suffix [-m:æ-] is from [-mpæ-] by consonant gradation.

- (95) Suomessa onkin tapana – kuten muinaisissa ritarilinnoissa – valmistaa sauna
hyville ja tervetulleille vieraille.
suome-s:a on-kin tapa-na kuten muinais-i-s:a ritari-lin:o-i-s:a
Finland-INE be[3sg]-also custom-ESS as ancient-PL-INE knight-castle-PL-INE
valmista-: sauna hyv-i-l:e ja terve-tule-i-l:e viera-i-l:e
prepare-INF sauna[acc] good-PL-ALL and well-come-PL-ALL guest-PL-ALL
'In Finland it is also the custom, just as in ancient castles of the knights, to
prepare a sauna for good and welcome guests.'
[sauna]: In the singular, the accusative case may have the same form as the nominative
singular under certain conditions.
- (96) Sauna tarjotaan yhtä hyvästä sydäimestä vieraitten nautintoa silmälläpitäen
sauna tarjo-ta:n yh-tæ hyvæ-stæ sydæme-stæ viera-i-t:en
sauna[acc] offer-IMPRS one-PTV good-ELA heart-ELA guest-PL-GEN
nautinto-a silmæ-l:æ-pitæ-e-n
enjoyment-PTV eye-ADE-keep-while-INS
'Sauna is offered with the aim of providing enjoyment to one's guests just as cordially'
- (97) kuin mitkä muut herkut tahansa
kuin mitkæ mu:-t herku-t tahansa
as any[PL.ACC] other-PL.ACC treat-PL.ACC whatsoever
'as any other treats whatsoever'
- (98) ja vastaanotetaan kiitollisuudella.
ja vasta:note-ta:n ki:to-lisu:de-l:a
and accept-IMPRS gratitude-expression-ADE
'and one accepts it with an expression of gratitude'
- (99) Sauna on ja tulee varmaan olemaan Suomessa vieraileviin ulkomaalaisiin nähden
parhaita mainosvälineitä.
sauna on ja tule-: varma:n ole-ma:-n suome-s:a
sauna be[3sg] and come-3SG certainly be-GER-ILL Finland-INE
vieraile-v-i-i:n ulko-ma:-lais-i-i:n näh-de-n
visit-PRS.ACT.PTCP-PL-ILL outside-country-inhabitant-PL-ILL see-while-INS
parha-i-ta mainos-væline-i-tæ
best-PL-PTV advertising-means-PL-PTV
'Sauna is and will certainly continue to be one of the best means of advertising for
foreigners visiting in Finland.'
[tule:]: Finnish does not have a specific future tense, but the future is sometimes marked by
[tule:] 'come', used as an auxiliary verb.
[næhden]: This word, when used as a kind of a postposition ('concerning'), governs the
illative case.

3.7 Exercises

3.7.1 RUSSIAN

Analyze the following Russian passage (Tolstoy 1852: Chapter 2, “Maman”) in the same way as with our example sentences and sample text. We give the passage in standard Russian orthography and pronunciation in IPA. We have also made a start on a word-by-word, morpheme-by-morpheme analysis and a free translation. But wherever we use italics, that is a placeholder for you to replace with a real analysis. We have supplied a few hints as to what information we expect – grammatical categories such as *case* and *number*, the word *gloss* when we expect an English word, *translation* for a free translation; but ? means we are leaving everything up to you. In many cases we have inserted morpheme boundaries “-” in the pronunciation, but when the morphemic structure is clear, we have sometimes left those out, for you to provide.

You should be able to draw all necessary information from the Sketch of Russian (§3.5) or by applying hints found elsewhere in this exercise.

- (1) Когда матушка улыбалась,
 ka'gda 'mat-uʃk-a uɫi'ba-ɫ-a-sʲ
 gloss mother-DIM-gender.case.number smile(IPFV)-tense-gender.number-?
 ‘translation,’
- (2) как ни хорошо было её лицо,
 'kak nʲi xara'ʃo 'biɫa i'jo ɫi'tso
 how ever gloss-gender.case.number gloss-?-? ?-? face-gender.case.number
 ‘no matter how pretty translation,’
 [xara'ʃo] is a predicate adjective with short endings.
- (3) оно делалось несравненно лучше,
 a'no 'dʲeɫa-ɫ-a-sʲ nʲi-srav'nʲenn-a ɫut-ʃi
 3SG-N.NOM make(IPFV)-?-?-? not-comparable-ADV ?-COMP
 ‘translation became translation,’
 Comparative adjectives do not inflect for gender, case, or number. [ɫut-] is the SUPPLETIVE stem for the comparative degree of [xara'ʃo]: not merely irregular, but based on a totally different root morpheme.
- (4) и кругом всё как будто веселело.
 i kru'gom ʲsʲo 'kag 'butta vʲisʲi'ɫe-ɫ-a
 gloss all_around gloss-??.? as though become_cheery(IPFV)-tense-?.?
 ‘translation seemed to translation.’

- (5) Если бы в тяжёлые минуты жизни

'jes'ʎi=bi f t'i'ʒot-iji m'i'nut-i 'ʒiz'n'i-i
 gloss=? ? hard-gender.case.number moment-F.ACC.number life-F.GEN.SG
 'translation'

Subjunctive particles can be rather distant from the verbs they apply to.

- (6) я хоть мельком мог видеть эту улыбку,

'ja 'xot' 'm'eʎkam 'mok 'v'id'i-tʃ
 person.case.number only briefly can[PST.M.SG] see(IPFV)-nonfinite
 'et-u u'ʎip-k-u
 this-F.case.number gloss-NOUN-gender.case.number
 'translation,'

- (7) я бы не знал, что такое горе.

'ja=bi n'i 'zna-ʎ 'ʃto ta'k-oja
 person.number=? ? know-tense.gender.number what such-gender.case.number
 'gor'i-a
 grief-N.case.number
 'translation what grief is.'

- (8) Мне кажется, что в одной улыбке состоит то, что называют красотой лица:

'mn'e 'kaʒ-it-sa ʃta v ad'n-oj
 .??.? seem(IPFV)-person.number.tense-? that gloss one-gender.LOC.number
 u'ʎipk-i sasta'j-it 't-o
 gloss-gender.case.number consist(IPFV)-person.number.tense that-gender.case.number
 'ʃto nazi'vaj-ut krasa't-oju
 which call(IPFV)-person.number.tense beauty-gender.case.number
 ʎi'ts-a
 gloss-gender.case.number

'translation:'

[-oju] is a more literary ending for the instrumental singular of paradigm 1. The object of [nazi'vatʃ] takes the instrumental case.

- (9) если улыбка прибавляет прелести лицу,

'jes'ʎi u'ʎipk-a pr'ibav'ʎaj-it 'pr'eʎ'ʎist'i-i
 gloss gloss-gender.case.number add(IPFV)-person.number.tense charm-F.ACC
 ʎi'ts-u
 gloss-gender.case.number

'translation,'

- (10) то лицо прекрасно;

'to li'tso pr'i'krasna
 then gloss-gender.case.number beautiful-gender.case.number
 'translation.'

- (11) если она не изменяет его,

'jes'ʃi a'n-a ni izmʃi'n'aj-it ji-'vo
 gloss 3SG-gender.case gloss change(IPFV)-person.number.tense 3SG-N.ACC
 'translation,'

- (12) то оно обыкновенно;

'to a'n-o abikna'vʃenn-a
 gloss person.number-gender.case average-gender.case.number
 '? an average one.'

- (13) если она портит его,

'jes'ʃi a'n-a 'portʃ-it ji-'vo
 gloss person.number-gender.case spoil(IPFV)-person.number.tense ?
 '?'

- (14) то оно дурно.

'to a'no 'durn-a
 gloss parse ugly-gender.case.number
 '?'

3.7.2 LOANWORDS IN FINNISH

Examine carefully the following loanwords from various European languages in Finnish and formulate as many general observations as you can about the way Finnish adapts foreign words to the Finnish phonological system. Take note, for example, of what regular sound substitutions occur and what constraints Finnish imposes on foreign words in terms of syllable structure and what sounds can occur in what positions. For the purposes of this exercise you may assume that, unless otherwise indicated, foreign originals are pronounced roughly as in English. For example, you may assume that in the language from which Finnish borrowed the word [pank:i] 'bank', this word, as in English, also began with [b] and ended in [k]. Ignore unsystematic changes. Give a brief explanation for each type of adjustment that you discover in the data.

- a. [sohva] 'sofa'
- b. [lamp:u] 'lamp'
- c. [koulu] 'school'
- d. [silk:i] 'silk'
- e. [tukholma] 'Stockholm'
- f. [kahvi] 'coffee'
- g. [kapte:ni] 'captain'
- h. [lasi] 'glass'
- i. [pi:p:u] 'pipe'
- j. [tohtori] 'doctor'
- k. [tirehtø:ri] 'director'
- l. [kulta] 'gold'
- m. [pastori] 'pastor'
- n. [tik:u] 'stick'
- o. [kupari] 'copper'
- p. [musi:k:i] 'music'
- q. [ka:su] 'gas'
- r. [ru:su] 'rose'
- s. [tans:i] 'dance'
- t. [eŋ:lanti] 'England'
- u. [konjak:i] 'cognac'
- v. [pank:i] 'bank'
- w. [ranska] 'France' < Swedish [franska]
- x. [rouva:] 'married woman, Mrs.' < German [frau]
- y. [evankeliumi] 'gospel' < Latin [evangelium]
- z. [risti] 'cross' < Old Russian [krist]
- aa. [tavara] 'goods' < Russian [ta'var]

3.7.3 TURKISH VOWEL HARMONY RULES

Like Finnish, Turkish also has vowel harmony. There are some similarities and also some differences between the two languages. For example, Turkish does not have a neutral vowel category as far as vowel harmony is concerned. Examine the Turkish data given in the following table and (1) formulate the rules of Turkish vowel harmony and (2) compare and contrast these rules with the vowel harmony rules of Finnish that were given in the "Sketch of Finnish". Note that the Turkish vowel system contains the following vowels:

(1)

	Front		Back	
Height	Unround	Round	Unround	Round
High	i	y	ɯ	u
Non-high	e	ø	ɑ	o

As in Finnish, personal suffixes on nouns are possessive. Thus the first row means 'stone', 'my stone', 'stones', and 'my stones'.

(2) Turkish data:

Gloss	Stem	Stem-1SG	Stem-PL	Stem-PL-1SG
1. 'stone'	taş	taşım	taşlar	taşlarıım
2. 'face'	jyz	jyzım	jyzler	jyzlerim
3. 'donkey'	eşek	eşekim	eşekler	eşeklerim
4. 'banana'	muz	muzum	muzlar	muzlarıım
5. 'fish'	balık	balıkım	balıklar	balıklarıım
6. 'tooth'	diş	dişim	dişler	dişlerim
7. 'ear'	kulak	kulakım	kulaklar	kulaklarıım
8. 'flower'	çiçek	çiçekim	çiçekler	çiçeklerim
9. 'eye'	göz	gözüm	gözler	gözlerim
10. 'son'	oğul	oğlum	oğullar	oğullarıım
11. 'arm'	kol	kolum	kollar	kollarıım

3.8 Suggested readings

3.8.1 SHORT SKETCHES

- ✘ Comrie (2009), which was recommended in Chapter 1, is especially useful for Europe. Included are brief introductions to the Indo-European and Uralic families and their major branches.

3.8.2 INDO-EUROPEAN LANGUAGES

- ✘ *The Celtic languages* (Ball & Müller 2009). History and structure of Irish, Scottish Gaelic, Manx, Welsh, Breton, and Cornish.
- ✘ *The Germanic languages* (Harbert 2007). History and structure of North, West, and East Germanic, focused on syntax.
- ✘ *The Germanic languages* (König & van der Auwera 2002). Historical overview of Germanic, plus detailed sketches of over a dozen Germanic and German-based languages by experts.
- ✘ *The Indo-Aryan languages* (Cardona & Jain 2007). Overview of Indic history, sociolinguistics, and writing systems, plus detailed sketches of 20 languages and groups.
- ✘ *Indo-European language and culture: An introduction* (Fortson 2010). Excellent introduction to Indo-European linguistics, presenting key evidence used in the reconstruction of the protolanguage, with a separate chapter on each branch of Indo-European and a chapter devoted to Proto-Indo-European culture and archaeology.
- ✘ *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world* (Mallory & Adams 2006). Focuses on reconstructing Proto-Indo-European roots and provides a wealth of cultural information along the way.

- ❏ *The Romance languages* (Harris & Vincent 1990). Historical overview of Romance, plus detailed sketches of 10 Romance languages and groups; also, Romance-based creoles.
- ❏ *The Slavic languages* (Sussex & Cubberley 2011). Socio-historical evolution of Slavic and comparative phonology, morphology, and lexis of modern Slavic languages.

3.8.3 URALIC LANGUAGES

- ❏ *The Uralic languages* (Abondolo 1998). Detailed sketches by experts in the various branches cover the entire family.

3.8.4 LANGUAGES OF THE CAUCASUS

- ❏ Caucasian languages (Hewitt 1981). Sketches of characteristic phonology, morphology, and syntax of the different groups, plus brief texts from Georgian, Abkhaz, and Avar.

3.8.5 BASQUE

- ❏ *Standard Basque: A progressive grammar* (de Rijk 2007). Very thorough reference grammar.
- ❏ *The Basque language: A practical introduction* (King 1994). Still the standard Basque language textbook.

3.8.6 RUSSIAN

- ❏ *Russian: A linguistic introduction* (Cubberley 2002).

3.8.7 FINNISH

- ❏ *Finnish* (Whitney 1956). Most of the examples cited in the Finnish sketch were taken from this work.

Asia

Asia, our largest continent, encompasses nearly 30% of the earth's land mass and is home to nearly 30% of the world's languages and 60% of the world's population. Asia's most populous language family by far is Sino-Tibetan, with languages spoken by nearly 1.3 billion people. Across the world, only Indo-European has more speakers – an estimated 2.9 billion.

Table 4.1 summarizes the languages of Asia. This chapter begins with the Turkic language family, which includes Turkish, the westernmost of Asia's main languages. We then turn eastward to the languages of East Asia, Southeast Asia, South Asia, and Austronesia. Finally, we cover a few language families that are only peripherally present in Asia.

4.1 Altaic area

The Altaic area, named for the Altai mountains in Siberia near the Mongolian border, extends from Turkey to northeastern Siberia. This area includes the Turkic, Mongolic, and Tungusic language families (Figure 4.1). In addition, many

TABLE 4.1
Languages of Asia

Name	Section	Size	Location
Altaic area	§4.1	9	Central Asia
Paleosiberian area	§4.2	5	Russia
Sino-Tibetan family	§4.3	10	China
Hmong-Mien family	§4.4	7	China
Tai-Kadai family	§4.5	8	Thailand
Austroasiatic family	§4.6	9	Vietnam
Dravidian family	§4.7	9	India
Burushaski (BSK)	§4.8	5	Pakistan
Other languages	§4.9		



FIGURE 4.1 Some languages of North Asia. Families: ¹Turkic. ²Mongolic. ³Tungusic. ⁴Japonic. ⁵Chukotko-Kamchatkan.

people include the Japonic languages and Korean, and sometimes Ainu, in this area (Table 4.2). Altaic languages share striking similarities, including SOV order, agglutinative morphology, and usually vowel harmony systems. As it happens, these features also closely resemble the systems of many languages in the Uralic family.

There has been a long and contentious debate, going back to the mid-19th century, as to whether the Altaic languages are all genetically related. The alternative is that the similarities among them are due to centuries-long interactions among adjacent language groups. This question will be revisited at the end of this section. For now, we take the conservative stance of not calling the Altaic languages a *family*, because the genetic relationship hasn't been definitively proven.

4.1.1 TURKIC

The Turkic family comprises some 40 languages spoken over a wide area from Macedonia to the Russian Far East (Table 4.2, Figure 4.1). This vast coverage was the product of early medieval Turkish expansions from east Asia. This expansion replaced Indo-European languages with Turkic languages in much of Central and Western Asia. By the late 11th century, the Turkic language Turkish began to replace Greek in Anatolia – modern Turkey – and established a significant presence in Balkan Europe.

Turkic languages present a very compelling corrective to the idea that writing systems are inherent properties of languages. The earliest written records of a Turkic language, from the early eighth century, were found in the Orkhon Valley of Mongolia and were written in a rune-like script called *Old Turkic* or the *Orkhon script*. A variant of that script has been found as far west as Hungary, where, as a result of intense contact between the Turks and the Hungarians, it was borrowed to represent the Uralic language of that country.

Turkic languages came to adapt a variety of other scripts. The story of writing in Turkmen is like that of many Turkic languages in central Asia (Clement 2008). Originally, it was written in Arabic script, familiar to all Muslims as the script in which the Koran is written. In the 1920s, Turkmenistan, as part of the Soviet Union, was subject to the union's LATINIZATION campaign, by which Turkmen, along with dozens of other languages, was furnished with a Latin writing system. But by 1940, the Soviet latinization campaign was replaced with a CYRILLICIZATION campaign, and Turkmen came to be written with a Cyrillic alphabet there. Soon after Turkmenistan became independent with the dissolution of the Soviet Union, it was furnished with a new Latin writing system, to highlight its new freedom from Russian cultural and economic influence. But many people continued to use the Cyrillic script as a way of expressing opposition to the totalitarian policies of the Turkmen president, who promulgated the new Latin writing system. And all this time, Turkmen speakers in Afghanistan, which was never part of the Soviet Union, have continued to write Turkmen in Arabic script.

TABLE 4.2

Some languages of the Altaic area

Name	Size	Location
Turkic	9	Central Asia §4.1.1
· Southwestern Turkic		
· · Turkish (TUR)	8	Turkey
· · Azerbaijani (AZB)	8	Iran
· · Crimean Tatar (CRH)	6	Ukraine
· · Turkmen (TUK)	7	Turkmenistan
· · Salar (SLR)	5	China
· · Gagauz (GAG)	6	Moldova
· Northwestern Turkic		
· · Karachai-Balkar (KRC)	6	Russia
· · Kumyk (KUM)	6	Russia
· · Bashkir (BAK)	7	Russia
· · Volga Tatar (TAT)	7	Russia
· · Karakalpak (KAA)	6	Uzbekistan
· · Kazakh (KAZ)	7	Kazakhstan
· · Kyrgyz (KIR)	7	Kyrgyzstan
· Southeastern Turkic		
· · Uyghur (UIG)	7	China
· · Uzbek (UZN)	8	Uzbekistan
· Northeastern		
· · Altai (ALT)	5	Russia
· · Tuvan (TYV)	6	Russia
· · Yakut (SAH)	6	Russia
· Chuvash (CHV)	7	Russia
Mongolic	7	Mongolia §4.1.2
· Dagur (DTA)	5	China
· Southern		
· · Tu (MJG)	6	China
· Central		
· · Mongolian (MON)	7	Mongolia
· · Buriat (BUA)	6	Russia
· · Oirat (XAL)	6	Russia
· Mogholi (MHJ)	0	Afghanistan
Tungusic	5	China §4.1.3
· Xibe (SJO)	5	China
· Manchu (MNC)	2	China
· Nanai (GLD)	4	Russia
· Evenki (EVN)	5	China
· Even (EVE)	4	Russia
Korean (KOR)	8	Korea §4.1.4
Japonic	9	Japan §4.1.5
· Japanese (JPN)	9	Japan
· Ryukyuan		
· · Okinawan (RYU)	6	Japan
Ainu (AIN)	2	Japan §4.1.6

Today, Turkish has about three times as many speakers as any other Turkic language. We have already discussed Turkish as an excellent example of the agglutinative morphological type in Chapter 1 (4). As an exercise in Chapter 3 (§3.7.3), we learned that Turkish has a vowel harmony very similar to that of Finnish and some other Uralic languages (§3.2). Despite the vast geographic expanse they cover, Turkic languages are for the most part still very similar to one another, and there is a high degree of mutual intelligibility among many of them. Chuvash appears to be a bit of an outlier, though. For example, the Turkish plural suffix we encountered in exercise §3.7.3, [-lar] or [-ler] depending on vowel harmony, has essentially the same form in all Turkic languages except Chuvash, which has [-sem]. Such differences have led some scholars to conclude that it may form one branch of Turkic on its own (along with some extinct languages), with all the other languages forming a separate branch (Róna-Tas 2007). Nevertheless, even Chuvash agrees with the other Turkic languages in having vowel harmony, agglutination, and SOV word order.

4.1.2 MONGOLIC

The Mongolic language family comprises languages spoken by perhaps seven million people (Table 4.2, Figure 4.1). Most of the Mongolic languages are spoken in Mongolia and in adjacent parts of Russia and China, especially Inner Mongolia. A dialect of Oirat is also spoken in Kalmykia, a Russian republic by the Caspian Sea.

By far the dominant language in the group is Mongolian. Mongolian uses two writing systems (1). One is based on the traditional Mongolian script (1a), which ultimately traces its history to Aramaic, the right-to-left script that spawned many other Asian scripts, including Arabic and Hebrew. Its immediate ancestor is a script that was used for Uyghur, a Turkic language spoken in China (§4.1.1). Apparently as an adaptation to the top-down writing direction of Chinese, that script was rotated 90 degrees counterclockwise, and Mongolian retains this vertical orientation. The more recent writing system is based on Cyrillic (1b; §3.5.2) and was introduced in Mongolia, then a Soviet satellite, during the Soviet Union's cyrillicization project. The traditional Mongolian script remained in use in China. Now that Mongolia is independent, there is some interest in reviving the traditional Mongolian script there. Our oldest substantial written record of Mongolian is the *Secret history of the Mongols* from the time of Genghis Khan, around 1240.

(1) [mɔŋɣɔɮ] 'Mongol'

a. Mongolian script:

The image shows the Mongolian script for the word 'Mongol'. It is a vertical script written from right to left. The characters are stylized and connected, forming a single vertical column of text.

b. Cyrillic script: Монгол

Proto-Mongolic resembled the Uralic and Turkic languages in being an agglutinative, suffixing language with SOV word order and vowel harmony (Nugteren

2011). In general, these facts apply to the currently spoken Mongolic languages as well, although some no longer exhibit vowel harmony.

4.1.3 TUNGUSIC

Most of the Tungusic languages are spoken in eastern Siberia and northeastern China (Table 4.2, Figure 4.1). The most famous representative of this family is Manchu, the language of the people who founded the Qīng Dynasty, which ruled China until 1912. Manchu is an agglutinative language with SOV word order, vowel harmony, and about half a dozen different cases. It is written in the vertical Manchu script, which was derived from the Mongolian script (§4.1.2). As early as the 18th century, Manchu speakers began to assimilate to the dominant Han culture, and their ethnic language gave way to Chinese. Today perhaps only a dozen or so people still speak Manchu natively. However, in western China some 30,000 people speak Xibe, a language closely related to Manchu. They descend from members of a military garrison established by a Qīng emperor in the 18th century. Today some Chinese Manchus, eager to revive their ethnic language, consider Xibe speakers to be among the most reliable sources of inspiration for what colloquial Manchu should sound like (I. Johnson 2009). But all Tungusic languages, Xibe included, are under intense competition from the national language, Mandarin Chinese or Russian, which almost everybody speaks.

4.1.4 KOREAN

Korean (Table 4.2, Figure 4.1) is the national language of both North and South Korea, which have been divided politically since 1945. Korean is agglutinative and SOV. Today it has only traces of vowel harmony, but in the Middle Korean period there was a strong vowel harmony system. Like its neighbor Japanese, Korean grammar includes a complex, pervasive set of HONORIFIC forms, which reflect the status of a speaker in relation to the person spoken to or about. Korean has a contrast between plain (e.g. [p]), aspirated [p^h], and TENSE [p̚] consonants. The tense consonants developed from consonant clusters in Middle Korean and are a subject of phonetic and phonological controversy. They have been variously reported as fortis, long, faucalized, forced, glottalized, having greater air pressure and articulatory tension than plain consonants, and as causing a higher pitch in the following vowel; others suggest tense stops sound just like ordinary voiceless stops in familiar languages such as Spanish (Chang 2006).

The Korean writing system, which is called *hangeul* in South Korea, was invented in 1443 as a way of opening up literacy to people who did not have the leisure to learn the thousands of Chinese characters that were used to write Korean. It is an abugida (§2.1.3.4), with all vowels indicated by a diacritic written to the right or below the onset consonant of each syllable. For example, [mikuk salam] (phonetically, [miguksalam]) is written by phonemograms clustering into

four compound symbols, one for each syllable (2a), but one can easily identify the individual symbols for each of the phonemes (2b):

(2) ‘American’ in Korean hangeul

a. Actual orthography

미 국 사 람
mi kuk sa lam

b. Letter by letter

ㅁ | ㄱ ㅈ ㅊ ㅌ ㄹ ㅍ ㅍ ㅁ
m i k u k s a l a m

Hangeul is often described as a FEATURAL SCRIPT because its individual letters give clues as to a consonant’s place and manner of articulation. Display (3) shows how strongly the shapes of the Korean letters correlate with their pronunciation. For example, a right angle in the lower left corner indicates a nonsibilant alveolar; an added horizontal line turns the pronunciation into a stop; yet another horizontal line indicates aspiration. The system isn’t perfect, but does show a remarkable amount of phonetic sophistication tempered by the need to keep the individual symbols relatively simple and easy to read and write.

(3) Korean letter shape patterns (Treiman & Kessler 2014: 35)

ㅁ m	ㄴ n	ㄷ s	ㅇ ɲ
ㅂ p	ㅌ t	ㅈ tɕ	ㅋ k
ㅍ p ^h	ㅍ t ^h	ㅊ tɕ ^h	ㅋ k ^h ㅎ h

4.1.5 JAPONIC

Japanese is by far this family’s dominant member (Table 4.2, Figure 4.1). It is an agglutinative SOV language without vowel harmony. Japanese is well known for its complex writing system. It also has a relatively intricate system for marking sociolinguistic REGISTERS, styles of language that vary depending on the social situation. One dimension is a distinction between plain language, which is appropriate when talking with friends or when writing for a general audience, and polite language, which is appropriate for talking with other people. The most easily noticed characteristic of polite language is the suffixation of [-mas-] to the main verb of the sentence. The other important sociolinguistic dimension is the amount of respect the speaker needs to express toward the people and things mentioned. There are many special words and expressions that are preferred when referring to one’s social superiors and the people and things or actions associated with them; these special forms are called *honorifics*. Honorifics are typically extended to members of groups that the addressee is a member of. When referring to oneself or the people and things associated

with one's group, one must use plain (nonhonorific) or even humble, HUMILIFIC, words or phrases. These rules can lead to seemingly paradoxical outcomes. For example, an employee would be careful to use an honorific when addressing his boss, or talking about his boss with fellow employees. But when talking about his boss with somebody outside the company, such as a customer, he would not use an honorific; that would sound as if he were arrogantly exalting his own group.

You may have seen Japanese featured on old lists of language isolates. At one time, Ryukyuan languages such as Okinawan were treated as dialects of Japanese. In fact, the Ryukyuan languages have been diverging from Japanese, and from each other, for more than 1,000 years, and they are not mutually intelligible. Now that people have agreed that they should be considered separate languages, Japanese is no longer considered an isolate but a member of a family with perhaps a dozen members. The Ryukyuan languages have SOV word order, but differ from Japanese in many respects. Particularly noteworthy is the presence of glottalized and SYLLABIC fricatives in several Ryukyuan languages. A syllabic consonant is one that functions as a syllable nucleus, filling the role usually filled by vowels. For example, [tʃkʃɕu] means 'moonlit night' in the Shimozato dialect of Miyako (MVI) (Karimata 2015: 120). It is a three-syllable word, with the two [s] sounds forming the nucleus of their syllables.

4.1.6 AINU

This language isolate (Table 4.2, Figure 4.1) is spoken by just a handful of speakers, although around 15,000 individuals identify themselves as ethnically Ainu. The basic word order of Ainu is SOV, and the phoneme inventory is similar to that of Japanese. But in many respects the two languages are very dissimilar. Ainu nouns can take possessive suffixes, like Turkish. Unlike typical languages in the Altaic area, Ainu has prefixes and noun incorporation. Example (4a) is an ordinary sentence with a transitive verb and a separate direct object: the noun [inau], a wooden prayer stick. The verb has a prefix that says that it is transitive and has a first person singular subject; an explicit subject pronoun is not required. In (4b), the direct object has been incorporated within the verb. Because that verb can take no further object, in this version of the sentence it takes a first person singular subject agreement marker appropriate for intransitive verbs. Verbs can incorporate other elements as well, such as adverbs, resulting in very long words that invite characterizing Ainu as polysynthetic.

(4) Ainu 'I make an inau.' (Shibatani 1990)

- a. inau a-ke
inau 1SG.TR-make
- b. inau-ke-an
inau-make-1SG.INTR

4.1.7 CLASSIFICATION

There is no doubt that Turkic, Mongolic, Tungusic, Korean, Japonic, and Ainu are valid linguistic clades. What is controversial is whether there are any proven relations between any of these clades. A very long-lived theory, the Altaic theory, has held that Turkic, Mongolic, and Tungusic are demonstrably related to each other; many include Korean and Japonic in that set, and a few include Ainu.

The Altaic theory was inspired by the compelling fact that these languages look very similar to each other, especially when contrasted with the Indo-European languages that were more familiar to early European linguists. They place the verb at the end of the sentence, they have vowel harmony, they are agglutinative, they use no prefixes, they use postpositions instead of prepositions, and they do not have gender, to cite some of the more salient features. This impression is reinforced by the fact that many words are shared across the language families.

However, such similarities cannot truly prove that the languages are related, for two main reasons. First, far from being bizarre phenomena crying out for an explanation, many of the seemingly exotic features of the Altaic languages are very common across the languages of the world. SOV word order is at least as common as SVO. SOV languages tend to have postpositions, so it is an exaggeration to count SOV order and postpositions as two separate pieces of evidence. Prefixes are much less common than suffixes in most parts of the world, most language families do not have Indo-European-style gender systems, and even vowel harmony is quite common. Sharing any or all of these features could be due to chance, and so cannot be used to prove that the languages are related. After all, most of the properties that were cited as characteristic of Altaic languages are also characteristic of Uralic languages, but nowadays almost everyone rejects Ural-Altaic as a valid language clade. The other reason the shared features of the Altaic languages do not prove they are related is that the speakers of Mongolic languages had intensive contact with speakers of Turkic and Tungusic languages. Not only vocabulary but also other features such as word order can easily spread between languages when there are many bilingual speakers.

Many of these early misconceptions have been addressed in more recent work, which has put more emphasis on recurrent sound correspondences and shared morphological paradigms. S. Starostin and colleagues have gone so far as to publish a three-volume *Etymological dictionary of the Altaic languages* (2003), which presents thousands of cognate sets and proposes sound correspondences and reconstructions covering the core Altaic languages as well as Japonic and Korean. But even this monumental work, which used standard techniques such as the comparative method, has not been immune to criticism, including an extended review by a prominent proponent of the Altaic theory who said that the dictionary actually caused him to change his mind and reject the Altaic theory (Vovin 2005). When correspondences are sought across the entire vocabularies of a large number of languages, without any prior constraints on how words have to match morphologically or semantically before they can be included in the same

correspondence set, it is exceedingly difficult to evaluate whether the mass of evidence proves the languages are related or is no more than one would expect to find by chance among unrelated languages.

In this book, we have adopted the standard that a group of languages should be treated as related to each other only if a genetic relationship has been proved to the satisfaction of the community of linguists. If we were to hazard a guess, it would be that most historical linguists believe that a genetic relationship among at least Mongolic, Tungusic, and Turkic is plausible or even very likely; but many of them would hesitate to say that the relationship has been proved. Significantly higher numbers would be skeptical that Japonic and Korean are related to the other languages, and even more would be skeptical about Ainu, whose morphology is typologically quite different from that of the other languages. We therefore stop short of referring to Altaic as a family in this book, but rather treat it as a nongenetic grouping of families.

There are many other proposals for relatedness among the Altaic languages. Best known is the *NOSTRATIC* hypothesis, which holds that the core Altaic languages are not only related to each other, but also to the Indo-European and Uralic languages. Most Nostraticists also include other language families, such as Afro-Asiatic (§5.1), Kartvelian, and Eskimo-Aleut (§7.1). Salmons & Joseph (1997) have published a useful collection of papers from a conference where linguists debated the quality of the evidence supporting this far-reaching proposal.

4.2 Paleosiberian area

Paleosiberian (Table 4.3, Figure 4.1) is a geographical grouping of language families and isolates spoken in Siberia. The *paleo-* in the name reflects the notion that

TABLE 4.3

Some languages of the Paleosiberian area

Name	Size	Location
Chukotko-Kamchatkan family	4	Far East
· Chukchi (ЧКТ)	4	Chukotka
· Koryak (КРҮ)	4	Kamchatka
· Alutor (АЛР)	2	Kamchatka
· Itelmen (ИТЛ)	2	Kamchatka
Ket (КЕТ)	3	Krasnoyarsk
Yukaghir family	2	Sakha
· Northern Yukaghir (ҮКГ)	2	Sakha
· Southern Yukaghir (ҮУХ)	2	Sakha
Nivkh (НИҮ)	3	Sakhalin

Note: Krasnoyarsk is in the Siberian Federal District; the other locations are in the Far Eastern district.

these descend from ancient languages that predated other languages of the region belonging to the Turkic and Tungusic groups. In addition to the languages listed in the table, Ainu is also sometimes included in this grouping.

According to Comrie (1981: 238), the Paleosiberian languages tend to be agglutinative, and most exhibit SOV word order, but on the whole they are not particularly similar to one another. Ket, for example, unlike the rest of the Paleosiberian languages, has several phonemic tones, *INFIXES* (affixes that are inserted inside a root, splitting it in two), and few nonfinite verb forms. Nivkh has a well-developed system of numeric *CLASSIFIERS*: numbers must be followed by one of 26 different morphemes depending on the identity or class of the counted object (see §4.10.3.3 for a discussion of classifiers in Chinese). Chukchi and Koryak are ergative languages and have vowel harmony.

Ket, the last survivor of a Yeniseian language family, is by far the westernmost representative of this group. Despite its being some 5,000 kilometers from North America, for the past 90 years linguists have been proposing that it is related to the Na-Dene languages, a language family spoken in the western parts of Canada and the United States (§7.2). The inspiration for this hypothesis was that Yeniseian languages have long strings of agglutinative prefixing on verbs, a typological feature that is very rare in Asia but also encountered in Na-Dene. This connection invites the exciting conclusion that it is possible to reconstruct a Proto-Dene-Yeniseian, which was presumably spoken in Siberia many millennia ago and has descendants in both the old and new worlds. The most recent attempt to link the two was by Vajda (2011), who based his demonstration not only on comparisons of the verb prefixes but also on sound correspondences drawn from 110 sets of proposed cognates. Vajda's work has been well received by many linguists, but others deny that there is enough evidence to consider the connection to be definitively proven (L. Campbell 2011; G. Starostin 2012; both of which include responses by Vajda).

4.3 Sino-Tibetan

There is widespread agreement that the Sino-Tibetan family (labeled with “1” in Figure 4.2) has two branches, Sinitic and Tibeto-Burman, but much remains to be uncovered about the internal specifics of each branch.

4.3.1 SINITIC

The Sinitic branch contains what has traditionally been described as Chinese (ZH0) and its dialects (Table 4.4, Figure 4.2). The description of the different clades as *dialects* is understandable because they are all spoken in the same country, and all but Mǐn diverged fairly recently in Chinese history, having Middle Chinese as their parent. Furthermore, the written language has always been rather uniform throughout the country. But the spoken forms of Chinese have diverged so much from each

TABLE 4.4

Some Sinitic languages

Name	Size	Location
Sinitic	10	China
· Mandarin Chinese (CMN)	9	Northwest, interior
· · Northern		Hénán to Hēilóngjiāng
· · Northwestern		Qīnghǎi to Shānxī
· · Southwestern		Yúnnán to Húběi
· · Eastern		Ānhuī, Jiāngsū
· Wú (WU)	8	
· · Northern Wú		southern Jiāngsū
· · Southern Wú		Zhèjiāng
· Xiāng (HSN)	8	Húnán
· Yuè (YUE)	8	Guǎngxī to Hong Kong
· Hakka (HAK)	8	Guǎngdōng
· Gàn (GAN)	8	Jiāngxī
· Mǐn	8	Fújiàn
· · Central Mǐn (CZO)	7	Yǒng'ǎn
· · Eastern Mǐn (CDO)	7	Fùzhōu
· · Northern Mǐn (MNP)	8	Nánpíng
· · Southern Mǐn (NAN)	8	Xiàmén, Táiwān

other that they are by no means mutually intelligible, and it is becoming more common, particularly in western linguistics, to describe them as separate languages.

Mandarin is by far the most widespread and widely spoken of the Chinese dialects. It is called *Mandarin* after *mandarin*, a name for a high government official, from Malay [məntəri]. Chinese officials used this dialect to communicate with each other, since there was no assurance they could understand one another's native dialect. The choice of dialect was surely influenced by the fact that Nánjīng and Běijīng, China's capitals in recent centuries, were both in the Mandarin-speaking area. Confusingly, the name *Mandarin* now applies both to the standard accent of Chinese as well as to the fairly diverse dialect group to which that standard belongs. On the standard language, see §4.10. Mandarin dialects have changed word-final stops to a glottal stop, or, much more commonly, deleted them altogether; and final [m] has changed to [n]. Consequently, many morphemes that were distinguished in Middle Chinese became homophones in Mandarin dialects but remain distinct in other dialects.

The Wú language is spoken on the central coast. Its territory includes major cities such as Shànghǎi and Sūzhōu. The southern Wú dialects, spoken in Zhèjiāng, have conserved more of the distinctive Wú vocabulary, whereas the northern dialects have been greatly influenced by the neighboring Mandarin dialects in this respect. Most Wú dialects have preserved Middle Chinese voicing contrasts, so that, for example, they still have sounds like [b] and [z], which have been devoiced in most other dialects, including Mandarin.

The Xiāng language is often divided into Old and New Xiāng, mainly on the basis of which dialects are more conservative. The Old Xiāng dialects, in particular, retain voiced obstruents, like Wú, but the New Xiāng dialects have devoiced them, like Mandarin.

Yuè is spoken along the southern coast of China. The dialect of Guǎngzhōu is considered to be the most prestigious form of this dialect, and consequently the dialect as a whole is often called *Cantonese*, after the westernized name of that city. It is the predominant form of Chinese spoken in Hong Kong and Macau and in many overseas Chinese communities.

Hakka, which is also called *Kèjiā*, is spoken in southern China, Taiwan, and many overseas Chinese communities. It is often considered particularly similar to Gàn.

Mǐn is a very diverse group of dialects increasingly considered to be at least six separate languages. They are centered in Fújiàn and Táiwan. Southern Mǐn (Mǐnnán) is often referred to overseas as *Hokkien* dialects. Hokkien is widely spoken in Taiwan and in many overseas communities.

Sinitic languages are tonal, to a large extent monosyllabic, and have analytic morphology. Most of them have SVO word order with modifiers preceding the words modified. The earliest written records go back to about 1200 BC.

4.3.2 TIBETO-BURMAN

The Tibeto-Burman languages (Figure 4.2) are spoken in the general area of the Himalayas: Tibet and Yúnnán in Southwest China, Nepal, Bhutan, Burma, the “Seven Sisters” region of India, and parts of some adjacent countries such as Thailand. The languages have not yet been definitively classified for several reasons, ranging from practical matters of access to many parts of the area, to the sheer numbers of the languages, with estimates ranging from 250 to 450. Table 4.5 lists a small sampling of the Tibeto-Burman languages. It is arranged according to the classification of Matisoff (2003), but be aware that other sources often have very different classifications and widely differing names for the language groups as well as for the individual languages themselves.

By far the most prominent Tibeto-Burman language is Burmese, the official language of Myanmar. It is written in the Burmese script, as are other Tibeto-Burman languages of Myanmar. The Burmese script developed from Indian scripts about 1,000 years ago. Like most Indian scripts, it is an abugida, where every consonant letter is assumed to be followed by the vowel [ə] unless diacritics are added to indicate otherwise. The other half of the name of this language clade refers to Tibetan languages; one of them, Classical Tibetan (xCT), is the subject of our sketch later in this chapter (§4.11). It also uses an Indian-derived abugida, dating back to the seventh century. In contrast, Yi, a language spoken mostly in China, uses a script inspired by Chinese writing. The modern Yi script is a syllabography.

Many of the Tibeto-Burman languages have borrowed heavily from other languages, including Chinese. An extreme case is Bai, which appears to have about 70% Chinese vocabulary, a situation that leads some scholars to believe it actually

TABLE 4.5

Some Tibeto-Burman languages

Name	Size	Location
Kuki-Naga		India
· Mizo (LUS)	6	Mizoram
· Thado (TCZ)	6	Manipur
· Ao (NJO)	5	Assam
· Sema (NSM)	5	Nagaland
· Angami (NJM)	5	Nagaland
· Konyak (NBE)	5	Nagaland
· Tangkhul (NMF)	5	Manipur
· Lepcha (LEP)	5	Sikkim
Barish		India
· Bodo (BRX)	7	Assam
· Dimasa (DIS)	6	Assam
· Garo (GRT)	6	Meghalaya
Tani		India
· Miri (MRG)	6	Assam
· Adi (ADI)	5	Arunachal Pradesh
· Nyishi (NJZ)	5	Arunachal Pradesh
Himalayish		China
· Tibetan (BOD)	7	China
· Newari (NEW)	7	Nepal
Qiangic		China
· Southern Qiang (QXS)	6	Sichuān
· Tangut (TXG)	0	Níngxià
Jingpho-Asakian		Myanmar
· Jingpho (KAC)	6	Myanmar
Yi-Burmese		Myanmar
· Burmish		Myanmar
· · Burmese (MYA)	8	Myanmar
· · Atsi (ATB)	6	China
· · Lashi (LSI)	5	Myanmar
· · Maru (MHX)	5	Myanmar
· Loloish		China
· · Lisu (LIS)	6	Yúnnán
· · Yi (III)	7	Yúnnán
· · Akha (AHK)	6	Yúnnán
· · Lahu (LHU)	6	Yúnnán
· Naxi (NXQ)	6	Yúnnán
Bai (BCA)	7	China
Karenic		Myanmar
· Eastern Pwo (KJP)	7	Kayin State
· Paò (BLK)	6	Shan State
· S'gaw (KSW)	7	Ayeyarwady
· Red Karen (KYU)	6	Kayah State

belongs in the Sinitic branch of this family. Along with the Karenic languages, Bai is also special in having a Chinese-style SVO word order rather than the SOV word order found in other Sino-Tibetan languages.

Sino-Tibetan languages tend to be analytic, monosyllabic, and tonal. These examples from Yi, which are written in Yi script, are typical. In (5a), the verb is pronounced with a high tone. In (5b), exactly the same syllable, except with a mid tone, has a completely different meaning.

- (5) Yi
- a. 𑄎 [t^hi] ‘bite’
 - b. 𑄎 [t^hi] ‘exchange’

However, not all of the Sino-Tibetan languages exhibit these characteristics. For example, Classical Tibetan as well as many modern dialects of Tibetan spoken on the periphery of Tibet are not tonal languages, and tonal contrasts in Burmese and Jingpho may be a relatively recent secondary development. There is also a group of languages that are traditionally called PRONOMINALIZED LANGUAGES, whose verbs are inflected with affixes that indicate their subject and object. Some linguists believe that this feature is the result of diffusion from Munda languages (§4.6), whereas others feel that this feature is a very conservative one and should perhaps be reconstructed for Proto-Sino-Tibetan.

4.4 Hmong-Mien

This family used to be considered a branch of Sino-Tibetan – a view that still has adherents, especially in China. Clearly these languages have been greatly influenced by Chinese in many respects and over a long period of time. No doubt other influences come also from the neighboring Tai-Kadai languages. The traditional designation for Hmong-Mien uses the Mandarin name *Miáo-Yáo*. Up to 300 or 400 years ago, the languages of this family were centered in mountains of southern China, but speakers have since spread to nearby countries of Southeast Asia and beyond (see the languages labeled with “?” in Figure 4.2). There are two major branches. Our classification of the Hmongic languages (Table 4.6) is based on that of Ratliff (2010).

Here from Cao (1987) are common characteristics of both the Hmongic and Mienic branches:

- ❑ Consonants of all types, even nasals, have a phonemic contrast in voicing or aspiration; for example, [p] versus [p^h], [m] versus [m̥] (the latter spelled ⟨hm⟩ in the name *Hmong*).
- ❑ Syllable codas, when they exist at all, tend to be very simple in structure, permitting only a few types of consonants. For example,

TABLE 4.6

Some Hmong-Mien languages

Name	Size	Location
Hmongic	7	
· Pa-Hng (PHA)	5	China
· Chuanqiandian (CQD)		
· · Hmong (HMN)	7	
· · · Hmong Daw (MWW)	6	China
· · · Hmong Njua (HNJ)	6	Laos
· · · Hmong Don (HMF)		Vietnam
· · A-Hmao (HRM)	6	China
· Xiangxi Miao (MMR)	6	China
· Qiandong Miao (HMS)	7	China
Mienic	7	China
· Iu Mien (IUM)	6	China

all the languages permit nasal stops in the coda, but none permit fricatives.

- ❑ All of these languages have a large number of lexical tones, as many as 12 in some cases. Some Hmong languages are reported to have up to five LEVEL TONES – those that are characterized by a single pitch – in addition to contour tones such as rising and falling. The choice of tones sometimes depends on features of the onset consonant. For example, aspirated consonants are followed only by certain tones.
- ❑ All of them have fairly complex noun classifier systems that are used with numbers (see §4.10.3.3 for examples in Chinese).
- ❑ Modifying adjectives always follow the noun they modify, but classifiers and possessive pronouns precede the noun.
- ❑ Grammatical relations are signaled by word order, which is SVO, and by particles.

There are also sizeable differences between the two branches:

- ❑ Only Hmongic languages have uvular stops, such as [q] and [q^h].
- ❑ Mienic languages permit voiceless stops in syllable codas as well as nasals; Hmongic languages permit only nasals.
- ❑ Some Hmongic languages permit consonant clusters in syllable onsets, such as stop + LATERAL. A lateral consonant is one in which the air passes around the side of the tongue, like [l]. Thus Hmongic has clusters like [pl]. Hmongic also has consonants that are PRENASALIZED, beginning with a very brief period of nasalization, like [p^h] (as in Classical Tibetan, §4.11.2).

- ☒ Besides singular and plural, Hmongic languages also have a dual number, especially in the pronouns. That is, they may have different pronouns for referring to one thing, two things, and three or more things.
- ☒ A noun classifier may be used without a number in Hmongic.
- ☒ Demonstratives, nouns, verbs, and adjectives are placed after the head noun in Hmongic but before it in Mienic.
- ☒ Mienic languages have two sets of numbers, one native and one borrowed, as well as a distinctive set of ordinal numbers.

The similarities among the Hmong-Mien languages make this a highly coherent family, while the divergences point clearly to the two main branches.

4.5 Tai-Kadai

Like Hmong-Mien, Tai-Kadai was once regarded as a branch of Sino-Tibetan. Its languages (Table 4.7; see the languages labeled with “3” in Figure 4.2) have been very strongly influenced by Chinese. But most scholars, at least those outside China, now classify Tai-Kadai as an independent family, one that Chinese has probably borrowed from.

Speakers of Tai-Kadai number close to 100 million. Included in this family are Thai and Lao – the national languages of Thailand and Laos. Note that *Tai* is

TABLE 4.7
Some Tai-Kadai languages

Name	Size	Location
Tai		
· Northern	8	China
· · Zhuang (ZYB)	7	China
· Southwestern	8	Thailand
· · Thai (THA)	8	Thailand
· · Lao (LAO)	7	Laos
· · Shan (SHN)	7	Myanmar
· · Tai Dam (BLT)	6	Vietnam
Kam-Sui		
· Dong (KMC)	7	China
· Sui (SWI)	6	China
Kra		
· Qabiao (LAQ)	3	Vietnam
· Gelao (GIQ)	3	Vietnam
Hlai (LIC)		
· Hlai (LIC)	6	China
Ongbe (ONB)		
· Ongbe (ONB)	6	China

usually used as the name of the branch, whereas the heterographic homophone *Thai* is used to refer to the main language of Thailand. Both languages are in the Southwestern clade of the Tai branch, and they are highly similar, sharing more than half their vocabulary and much of their syntax. The sound systems of Thai and Lao are very close, marked by a three-way contrast in labial and alveolar stops – voiceless aspirated, voiceless unaspirated, and voiced – and a two-way contrast in velar stops – voiceless aspirated and voiceless unaspirated. Still, Thai and Lao are not mutually intelligible.

The basic word order in this family is generally SVO, and adjectives usually follow the nouns they modify. All of the Tai-Kadai languages are also tonal and tend to be monosyllabic and analytic. Tai-Kadai languages have been very strongly influenced by other languages, especially by Chinese. Chinese historical records show that Tai-Kadai-speaking peoples used to be located much farther north than they are found now, covering large portions of China south of the Yangtze River.

In a famous article published in *American Anthropologist*, Benedict (1942) proposed that the Tai-Kadai languages and the Austronesian languages (§6.1) are genetically related to each other. He called this expanded family *Austro-Tai*. But from the outset, most Austronesianists were dubious, and, as Thurgood (1993) has pointed out, the Tai-Kadai words with Austronesian cognates exhibit irregular correspondences, strongly suggesting that such items are simply loanwords from Austronesian into Tai-Kadai.

4.6 Austroasiatic

As suggested by the morpheme *austr* ‘south’, this family extends across south Asia, from India to Vietnam (Table 4.8; see the languages labeled with “4” in Figure 4.2). It has two major branches, Munda and Mon-Khmer, which are typologically very divergent from each other.

The Munda, or Western, clade is found in northeastern India. Its speakers are surrounded by Indo-European and Dravidian (§4.7) speakers, whose languages have greater social prestige. Munda languages appear to have been greatly influenced by their Indo-European and Dravidian neighbors. Typologically they are agglutinative, and very long sequences of affixes may be found, especially in verbs. Some languages, including Sora, have noun incorporation (discussed earlier with respect to Ainu, §4.1.6), which is usually a feature of polysynthetic languages. Nouns are divided into two major categories: animate and inanimate. Grammatical relations are signaled by word order, postpositions, and pronominal affixes. The basic word order is SOV. In general, these languages are typologically very different from their relatives in the Eastern branch of the family.

TABLE 4.8
Some Austroasiatic languages

Name	Size	Location
Munda	7	India
· North	7	India
· · Korku (KFQ)	6	Madhya Pradesh
· · Kherwari	7	Jharkhand
· · · Santali (SAT)	7	Jharkhand
· · · Mundari (UNR)	7	Jharkhand
· · · Ho (HOC)	7	Jharkhand
· South	6	Odisha
· · Juang (JUN)	5	Odisha
· · Sora (SRB)	6	Odisha
Mon-Khmer		
· Nicobarese	5	Nicobar Islands
· · Car (CAQ)	5	Car Island
· Khasian		India
· · Khasi (KHA)	6	Meghalaya
· Palaungic		Indochina
· · Parauk (PRK)	6	Myanmar
· Khmuic	6	Indochina
· · Khmu (KJG)	6	Laos
· · Khmer (KHM)	8	Cambodia
· Bahnaric		Indochina
· · Bahnar (BDQ)	5	Vietnam
· Katuic	7	Indochina
· · Kuy (KDT)	6	Thailand
· Vietic	8	Indochina
· · Vietnamese (VIE)	8	Vietnam
· Monic	6	Indochina
· · Mon (MNW)	6	Myanmar
· · Nyahkur (CBN)	4	Thailand
· Aslian	5	Thailand
· · Temiar (TEA)	5	Malaysia

The Mon-Khmer, or Eastern, clade has a much wider distribution than the Munda clade, with most of its languages in Indochina. Most prominent among them is Khmer, the national language of Cambodia, and Vietnamese. Vietnamese has borrowed massively from Chinese, especially during the T'ang dynasty, when Vietnam came under strong cultural and political influence from China. It was originally written with Chinese characters (like Korean and Japanese) until the seventeenth century, when Western Christian missionaries devised an orthography

based on the Latin alphabet. The languages of the Mon-Khmer group are mostly prefixing and tend to be monosyllabic. There is usually a large number of vowel phonemes. Some of the languages, notably Vietnamese, have developed tones, and some may be considered to be in a transition stage from nontonal to tonal languages. The most commonly found word order is SVO.

4.7 Dravidian

Dravidian, with well over 200 million speakers, has more speakers than any other family in Asia except Sino-Tibetan. Its history can be traced fairly far back, thanks to longstanding writing traditions in several of its languages.

Most Dravidian languages are spoken in south India and parts of Sri Lanka (Table 4.9; see the languages labeled with “5” in Figure 4.2). An exception is Brahui, which is spoken in Balochistan province in western Pakistan. The existence of this Dravidian pocket deep in the Indo-Iranian language area strongly suggests that Dravidian languages originally extended much farther north before being pushed to the south by the invading Indo-Iranian tribes (Indo-Europeans) in prehistoric times.

TABLE 4.9
Some Dravidian languages

Name	Size	Location
Southern		
· Tamil (TAM)	8	Tamil Nadu
· Malayalam (MAL)	8	Kerala
· Kannada (KAN)	8	Karnataka
· Kodava (KFA)	6	Karnataka
· Tulu (TCY)	7	Karnataka
· Toda (TCX)	4	Tamil Nadu
· Kota (KFE)	3	Tamil Nadu
South-Central		
· Telugu (TEL)	8	Andhra Pradesh
· Kui (KXU)	6	Odisha
· Gondi (GON)	7	Madhya Pradesh
Central		
· Duruwa (PCI)	5	Chhattisgarh
· Kolami (KFB)	6	Andhra Pradesh
· Naiki (NIT)	5	Andhra Pradesh
Northern		
· Brahui (BRH)	6	Pakistan
· Kurukh (KRU)	7	Odisha
· Malto (MJT)	5	Jharkhand

Dravidian experts are largely agreed on the subgrouping shown in Table 4.9, although it is not completely clear that the Northern group of languages actually forms a true clade with a Proto–North Dravidian ancestor. Brahui has been influenced to a great extent by Balochi, an Iranian language (§3.1.8). Other Dravidian languages, especially those that have a long literary tradition, have been to a large extent influenced by Indic languages (§3.1.8), especially Sanskrit (through the spread of Hinduism), and have in their turn influenced their Indic and Munda (§4.6) neighbors. For example, Dravidian languages going back to Proto-Dravidian have RETROFLEX CONSONANTS such as [ɻ], where the tongue curls back so that the tip is in the postalveolar or palatal region. Proto-Indo-European did not have such consonants, but they are found already in the oldest Sanskrit texts, due to diffusion from Dravidian.

Dravidian languages are mostly agglutinative in type although they do not exhibit very elaborate chains of affixes, as are found in such agglutinative languages as Turkish. Some of them have also developed a number of fusional traits. The word order is relatively fixed, usually SOV. One of the more unusual typological features in some Dravidian languages, such as Malayalam, is the presence of a three-way phonological contrast in place of articulation for CORONAL STOPS (those made with the front of the tongue): dental ([t̪], [ɲ̪]) versus alveolar ([t̬], [ɲ̬]) versus the aforementioned retroflex ([ɻ], [ɻ̃]). Due, no doubt, to the difficulty in making and perceiving such a fine distinction, very few languages have this three-way contrast; most of the languages merge the alveolar consonants with one of the other series.

4.8 Burushaski

Burushaski is a language isolate spoken in northern Pakistan by about 87,000 speakers (Table 4.1; BSK in Figure 4.1). The languages nearest to it are Indic, Iranian (§3.1.8), and Tibetan (§4.3.2), but it is very remote and isolated. Attempts to link it with languages of the Caucasus, Basque, and some other languages primarily on the basis of typological similarities remain very unconvincing.

Burushaski is an ergative language with an SOV word order. Its phonological inventory is very much like that of its Indic neighbors, except that it has uvular obstruents and does not have breathy-voiced stops. It is also reported to have at least three distinct tones.

All Burushaski nouns have one of four semantically determined genders:

- ♀ male humans
- ♀ female humans
- ♂ animals; fruits, parts of trees, objects made of wood, and some natural phenomena, especially heavenly objects such as the moon
- ♂ other inanimate objects

Verbs have suffixes agreeing with the gender of their subjects, and they have prefixes agreeing with the gender of their objects. This feature is very reminiscent of similar phenomena in the East Caucasian languages (§3.3.2) and is usually cited

by people supporting the hypothesis that Burushaski is genetically related to those languages.

In Burushaski, some nouns must appear with possessive prefixes. These are mostly nouns naming objects that are considered to be *INALIENABLY POSSESSED* by people and animals – things that do not normally exist independently of a possessor. In Burushaski, this includes parts of the body, the noun ‘name’, nouns denoting certain emotions, and so forth, although there are some strange items included as well, such as ‘stick’ and ‘pillow’.

4.9 Other languages in Asia

Languages of several other families are spoken in Asia but have a center of gravity in other geographical areas. For the sake of cladistic coherence, each of these languages is discussed with the rest of its family in the appropriate chapter.

The most important Austronesian languages (discussed in the chapter on Oceania, §6.1) spoken on the mainland of Southeast Asia are Malay and the Chamic languages (§6.1.2).

Modern languages belonging to the Afro-Asiatic family (discussed in the chapter on Africa, §5.1) and spoken in Asia are the Eastern Arabic dialects of the Arabian peninsula, Iraq, Syria, Jordan, and Lebanon; Hebrew in Israel; and Aramaic languages spoken primarily in Syria, Lebanon, and Iraq.

Most of the modern branches of the Indo-European family are centered in Europe, and therefore this family as a whole is discussed in the chapter on the languages of Europe (§3.1). The Asian languages belonging to the Indo-European family are found in Iran, Afghanistan, India, Pakistan, Bangladesh, and Sri Lanka; for the most part they belong to the Indo-Iranian branch of the family (§3.1.8). Asian representatives of the Uralic family belong to the Samoyedic branch of the family in northern Siberia (§3.2.1).

The languages of the Eskimo-Aleut family are primarily centered in Alaska, and therefore this family will be discussed in the chapter on the languages of the Americas (§7.1). As far as can be proved definitively, this is the only language family that is found in both Asia and North America, even though it is quite certain that all native languages of the Americas are ultimately descended from those languages that were brought by migrants from Asia who came across the Bering Strait. In the case of Eskimo-Aleut, the presence of this family on the Asian continent is due to remigration back to Asia from Alaska, and not to a group of speakers remaining behind in Asia after the migration of the rest to North America.

4.10 Sketch of Mandarin Chinese

4.10.1 GENETIC RELATIONSHIP AND GENERAL BACKGROUND

Chinese is a member of the Sinitic branch of the Sino-Tibetan language family. It is spoken by over 1 billion people in a large number of dialects in China,

Singapore, Malaysia, and Indonesia. There are also substantial emigrant communities in Vietnam, Thailand, Burma, and other parts of the world, including North America.

The standard official language is Pǔtōnghuà, the Mandarin of Běijīng shorn of its purely local Běijīng characteristics. It is the medium of instruction in both the People's Republic of China (PRC) and the Republic of China on Tái wān, though in some regions of the PRC, notably the Yuè-speaking areas, schools use the local dialect in addition to the standard language. The overseas Chinese communities, in most of which Mǐn and Yuè speakers form the majority, are also increasingly using standard Mandarin Chinese in their schools.

Although the written language is understood throughout China by educated, literate speakers regardless of their dialect, oral communication between speakers whose dialects belong to different dialect groups is often very difficult or even impossible. The dialectal differences are sometimes greater than those found, for example, between English and Dutch. Literacy in the standard written language does not give very reliable cues toward standard pronunciation, because it is essentially a logographic script. A Chinese character that has the same meaning in all the dialects may have a very different sound value from dialect to dialect.

The greatest degree of difference exists between Mandarin and the Mǐn language groups. It is only by tradition that people still refer to these two groups as being dialects of the same language rather than two different languages. The greatest linguistic diversity is found on the coast of China, whereas in the interior, where Mandarin dialects stretch from Myanmar all the way to Russia, there is still much mutual intelligibility among dialects.

The government of the PRC is in the process of carrying out a language engineering project on a vast scale. The basic objective of this project is to unify various regions of China by facilitating communication among speakers of different languages and different Chinese dialect groups. To accomplish this task the government is promoting language standardization and mandating the use of the standards in all government activities, including education, publishing, and broadcasting (China 2000). Pǔtōnghuà is the official language. The number of Chinese characters has been reduced somewhat, and a few thousand characters have been officially replaced by simplified variants that are easier to write. The last prong of the language standardization process has been to introduce pīnyīn, a way to write Chinese phonemically using Latin characters. The current idea is not that pīnyīn is to replace Chinese characters, but that it is a tool used to teach people how to speak Chinese in correct Pǔtōnghuà pronunciation. Tái wān and the special administrative regions of the PRC – Hong Kong and Macau – make less use of pīnyīn and use traditional characters rather than the simplified characters. Schoolchildren and adults in the PRC are now being taught the standard language, but a recent news release acknowledges that 30% of the population does not speak Mandarin – and many of the remaining 70% do not speak it well (Roberts 2014).

In Táiwan, the government has been successful in teaching everyone Mandarin, even older, illiterate people, but the government on Taiwan did not have to contend with the vast rural population of the Chinese mainland.

Chinese has greatly influenced the languages of its immediate neighbors through the overwhelming cultural prestige of the Chinese civilization and culture in East and Southeast Asia. Japanese, Korean, Vietnamese, and Thai have borrowed extensively from Chinese at different times. All of the mentioned languages, except Thai, have also borrowed the Chinese system of writing. Vietnamese, however, has abandoned it completely in favor of romanization introduced by Catholic missionaries, and Korean is in the process of abandoning Chinese characters, though they are still used in South Korea to a certain extent. In recent times, Chinese has borrowed some words from English and other European languages, but Chinese has resisted large-scale borrowings, such as what took place in Japanese.

The earliest written texts in Chinese go back to around 1300 BC and consist mostly of logographic inscriptions on oracle bones – bones and turtle shells used in pyromancy.

4.10.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

The majority of Chinese morphemes consist of a single syllable; this is why Chinese is often cited as an example of a monosyllabic language. This term is very often misunderstood to mean that such a language has only monosyllabic words, which certainly is not true of any Chinese dialect or, for that matter, any known language in the world. Although Mandarin roots are to a very large extent monosyllabic, it has a preponderance of disyllabic words and a smaller number of disyllabic morphemes, such as [pʷəʌxə] ‘mint’, some of which are loanwords from other languages. Also, the diminutive suffix [aɿʔ] fuses with the preceding syllable – that is, this suffix is phonetically less than one syllable in length. Be that as it may, the phonological system of any Chinese dialect is traditionally described in terms of syllable structure, which is essentially morpheme structure as well, rather than in terms of individual segments. Figure 4.3 summarizes the structure of the Mandarin Chinese syllable.

4.10.2.1 Consonants

Table 4.10 lists the consonants of Mandarin Chinese in IPA as well as in the official pinyin orthography. Note that Chinese has a contrast between plain and aspirated versions of all stops and affricates. English aspirates voiceless stops at the beginning of a stressed syllable, so a Chinese sequence like [p^hau] sounds like the English word *pow!* What is less commonly recognized is that the English stops [b], [d], and [g] tend to be fully devoiced, though unaspirated, at the beginning of an utterance. Thus a Chinese sequence like [pau] sounds like the English word *bough*. The pinyin spellings of the stops – ⟨b⟩ for [p] and ⟨p⟩ for [p^h], etc. – follows this English pattern.

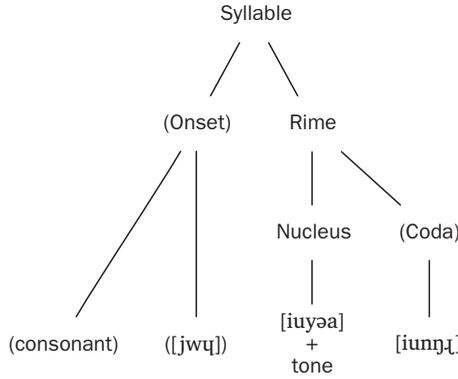


FIGURE 4.3 Mandarin Chinese syllable structure. Parentheses enclose optional elements. Square brackets enclose phonemes to choose from.

TABLE 4.10
Mandarin Chinese consonant phonemes

Manner	Labial		Alveolar		Postalveolar	Velar		
Stop								
Plain	p	⟨b⟩	t	⟨d⟩		k	⟨g⟩, ⟨j⟩	
Aspirated	p ^h	⟨p⟩	t ^h	⟨t⟩		k ^h	⟨k⟩, ⟨q⟩	
Affricate								
Plain			ʈʂ	⟨z⟩	ʈʃ	⟨zh⟩		
Aspirated			ʈʂ ^h	⟨c⟩	ʈʃ ^h	⟨ch⟩		
Fricative	f	⟨f⟩	s	⟨s⟩	ʃ	⟨sh⟩	x	⟨x⟩, ⟨ɣ⟩
Nasal	m	⟨m⟩	n	⟨n⟩		ŋ	⟨ng⟩	
Liquid			l	⟨l⟩	ɭ	⟨r⟩		

In addition to these consonants, the glides [j], [w], and [ɥ] also occur. They are the consonant forms of [i], [u], and [y], respectively, and pinyin spells them the same way as the vowels (§4.10.2.2). All these consonants and glides can occur in the onset of a syllable, with the exception of [ŋ]. They are severely restricted in the coda, however, as listed in Figure 4.3.

Before high front vowels ([i] and [y]), velar obstruents are palatalized to the ALVEOLOPALATAL region: the tongue blade makes a constriction against the alveolar ridge, and the body of the tongue is raised toward the palate. The velar stops become affricates: [k] is pronounced [tʃ] and [k^h] is pronounced [tʃ^h], while the fricative [x] remains a fricative: [ç]. These outcomes are comparable to the Russian palatalizations described in §3.5.3. In a departure from the phonemic principle, the pinyin orthography has special symbols for these allophones: ⟨j⟩, ⟨q⟩, and ⟨x⟩, respectively, but our IPA transcriptions, being phonemic, will not take note of the allophony.

position for the vowel as for the onset consonant, the vowel is treated as inherent in the consonant. After postalveolar sounds, including [ʃ], a similar allophone is heard for [i], although the more posterior pronunciation may suggest a slightly different representation such as [ɿ]; the unofficial symbol “ɿ”, which suggests a retroflex “ɿ”, is often used by Sinologists. Because we generally use phonemic notation, these phones will be notated as [i], which happens to agree with pīnyīn.

Table 4.11 shows how vowel phonemes are pronounced in different environments. As an example of how to read the table, the row labelled “j __ u” lists cases where the vowel in the column header is preceded by [j] and followed by [u]. So under column “ə”, you can see that the phonemic sequence [jəu] is pronounced [jəu]. Many apparent gaps – sequences that do not exist – are due to constraints that require contrast among phones within a morpheme. Glides, which are essentially high vowels, do not occur next to other high vowels, with the one exception of [juŋ]; nor can a glide and its corresponding vowel appear together in the same rime (e.g. there is no * [jai] or * [wau]).

Allophony is widest in the case of [ə], so much so that the pīnyīn orthography uses a variety of symbols for it: most notably, pīnyīn spells the phoneme as ⟨o⟩ when it assimilates to the rounding of an adjacent phoneme, and as ⟨e⟩ when it does not. Pīnyīn also omits the vowel letter ⟨e⟩ entirely when it would be written after another vowel letter and no ambiguity would occur; these cases are indicated with parentheses in the table. This omission is partly an orthographic convenience, but also reflects the fact that many speakers reduce or even drop the vowel in such environments.

One peculiarity of pīnyīn is that syllables are not allowed to begin with the letters ⟨i⟩ or ⟨u⟩. In cases where Table 4.11 suggests that should happen, ⟨i⟩ is converted to ⟨y⟩ if it is followed by a vowel; otherwise, ⟨y⟩ is prepended to the ⟨i⟩. For example, [in] ‘sound’ is spelled ⟨yīn⟩, and [jəu] ‘right’ is spelled ⟨yòu⟩ – note that the letters that are parenthesized in Table 4.11 have to be spelled in this context. Syllable-initial ⟨u⟩ changes to or adds ⟨w⟩ analogously. These changes make it easier for the reader to parse polysyllabic words: for example, the word ⟨pīnyīn⟩ breaks into the syllables [p^hin̩.in̩]; without the ⟨y⟩, it could be read as * [p^hi̯n̩.nin̩].

4.10.2.3 Tones

Mandarin dialects have fewer tones than most of the non-Mandarin dialects spoken in South China. Standard Mandarin has four contrastive lexical tones, traditionally numbered as shown in Table 4.12.

For most languages, IPA represents tone by means of diacritic marks placed over the vowel; for example, in [tán] the vowel is pronounced with a high tone. Such a system would be workable for Mandarin, but for most Chinese dialects there are so many different tones that the IPA diacritics would become unwieldy. This problem was addressed by the Chinese linguist Y. R. Chao, who devised a system of TONE LETTERS (Chao 1930). These have the same general size as letters and are written on the same line with them, placed after all the letters in the syllable. The tone letters are iconic: the vertical line is a staff, and the lines that attach to it

any derivational affixes being added. For example, [xjaʋ] ‘down’ may function either as a verb or as a kind of locative postposition:

- (6) a. wəʋ xjaʋ ʃanʅ
 1SG down mountain
 ‘I descend the mountain.’
- b. ʃanʅ xjaʋ jəuʋ ɹənʅ
 mountain down exist person
 ‘There are people below the mountain.’

4.10.3.1 Pronouns

The personal pronouns are listed in Table 4.13.

There are no gender distinctions of any kind in Chinese pronouns. But because of the influence of European languages, in modern times the third person pronoun is written with different logograms depending on whether it refers to men (他), women (她), or things (它). But there is no impact on pronunciation or the rest of the language. In the spoken language, [tʰaʅ] is only rarely used to refer to inanimate things.

There is also a polite form of the second person pronoun, *nín* [ninʅ], which can be used both in the singular and plural.

Though the distinction is not very commonly observed in the standard language, Běijīng Mandarin has two different forms for the first person depending on whether or not the person addressed is being included. For those who make the contrast, [wəʋmən] is EXCLUSIVE (EXCL) and [tsaʅmən] is INCLUSIVE (INCL):

- (7) a. wəʋmən kʰyʋ kʰanʋ tʃənʋ-iŋʋ
 1.PL.EXCL go look electric-shadow
 ‘We are going to the movies.’
 I and at least one other third party, not including you.
- b. tsaʅmən kʰyʋ kʰanʋ tʃənʋ-iŋʋ
 1.PL.INCL go look electric-shadow
 ‘We are going to the movies.’
 You are included.

TABLE 4.13
**Personal pronouns
 of Mandarin Chinese**

Person	SG	PL
1	wəʋ	wəʋmən
2	niʋ	niʋmən
3	tʰaʅ	tʰaʅmən

4.10.3.2 Nouns and nominal constructions

Although Mandarin pronouns regularly distinguish between singular and plural number, nouns do not show such a distinction. Thus, [ɣən˥] can mean either ‘person’ or ‘persons’ depending on the context. When it is necessary to indicate plurality unambiguously, modifiers such as ‘many’ are used. In addition, the plural suffix [-mən], which occurs in the pronouns, may be added to nouns denoting people.

4.10.3.3 Noun classifiers

There is no gender system as in Russian, French, and many other European languages, but nouns are grouped into classes of a different nature. Nouns of different classes are distinguished by the classifier (CLF) used with them when they are counted or when they are preceded by a determiner. Classifiers are similar to such English terms as *head* in *three head of cattle*. There are several dozen classifiers, which are chosen based on such factors as the appearance of the object being counted. The following examples illustrate only a few of these noun classifiers. Classifiers are not optional in the modern language; one cannot use numbers or certain other types of determiners without attaching a classifier.

- (8) a. i˧ ɿ̃ tʃi˧ kʰjən˥pi˧
 one CLF pencil
 ‘one pencil’
 classifier for long, cylindrical objects
- b. san˧ wəi˧ kʰəi˧ɣən˥
 three CLF guest
 ‘three guests’
 polite classifier for people
- c. nəi˧ kə ɣən˥
 that CLF person
 ‘that person’
 ordinary classifier for people

The classifier [kə] is increasingly used as a general classifier. It can be used with virtually any noun, especially one for which there is no special classifier.

4.10.3.4 Noun compounding

Noun compounding is a very productive process in Mandarin Chinese. Indeed, the large majority of nouns in modern Chinese are compounds, typically formed from two elements, each of which were independent words in Old Chinese. The development of new technical vocabulary often exploits this capacity for compounding by building up new compounds that use native morphemes. Therefore,

Chinese technical terms are often very transparent morphologically, and one can often guess their meaning:

- (9) a. fəi¹-ki¹
fly-engine
'airplane'
- b. pʃən¹-ja¹-k^hi¹
change-pressure-implement
'transformer'
- c. fa¹-tʃən¹-ki¹
emit-electricity-engine
'electric generator'

4.10.3.5 Verbs and adjectives

Chinese verbs are not inflected for person and number. Adjectives, which in Chinese are a subclass of verbs, do not change to agree with the nouns they modify. Both verbs and adjectives can be followed by the particle [lə], which is a marker of the perfective aspect (recall the discussion of aspect from the Russian sketch, §3.5.4.5) and also functions as an inchoative, an indicator of a new state or situation. In other words, one must distinguish two different morphemes that happen to sound the same but have different functions.

- (10) a. t^hʃən¹ xəi¹ lə
sky black PFV
'The sky darkened.'
- b. wə¹ t^hi¹-fan¹ lə
1SG eat-rice PFV
'I ate' or 'I finished eating.'
- c. wə¹ pu¹ k^hy¹ lə
I NEG go INCHO
'I am no longer going to go.'
I was going to go, but changed my mind: a new situation.

Tense is marked not by inflection but by time words or by some other lexical indicators of time. A perfective marker usually implies that an action took place in the past, and future action is often implied by such verbs as [jau¹] 'want' or [xjaŋ¹] 'think':

- (11) a. miŋ¹t^hʃən¹ wə¹ k^hy¹ k^han¹ tʃən¹liŋ¹
tomorrow 1SG go look movie
'I shall go to see a movie tomorrow.'

- b. wəʌ jauʌ kʰyʌ kʰanʌ tʃənʌiŋʌ
 1SG want go look movie
 ‘I want to go see a movie.’

The verb *to be* has several functions in English, all of which are expressed by different verbs in Chinese. This state of affairs is not all that unusual in the world’s languages:

- (12) a. wəʌ ʃiʌ xʉəʌʃəŋ
 1SG be(COP) student
 ‘I am a student.’
 [ʃiʌ] is a COPULA – a linking verb.
- b. wəʌ ʔsaiʌ kjaʌ
 1SG be_located home
 ‘I am at home.’
 [ʔsaiʌ] shows location.
- c. jəuʌ ɿənʌ ʔsaiʌ mənʌkʰəuʌ
 EXIST person be_located door
 ‘There is a person at the door.’
 [jəuʌ] shows existence: it is an EXISTENTIAL VERB.

The same existential verb [jəuʌ] is used in possessive constructions such as the following:

- (13) wəʌ jəuʌ ljaŋʌ pənʌ ʃuʌ
 1SG EXIST two CLF book
 ‘I have two books.’

This example shows that, at least in some cases, what one might regard as the subject of the Chinese sentence is in reality the TOPIC of the sentence. Thus, the sentence may be translated more literally as ‘As for me, there exists a book’. In other words, ‘the book’ may be regarded as the subject of the verb ‘to exist’ because it is the book that does the existing, whereas ‘I’ is merely the topic about which the rest of the sentence supplies some information.

4.10.3.6 Resultative verb compounds

Some verbs can be combined directly with other verbs, such that the first verb indicates the main action while the second verb indicates the result of that action:

- (14) a. tʰaʌ xjaʌ-siʌ lə wəʌ
 3SG frighten-die PFV 1SG
 ‘He frightened me to death.’

- b. wəʌ k^hanʌ-kjənʌ t^haʌ
 1SG look-perceive 3SG
 'I see him.'

4.10.3.7 Negation

There are two negatives that precede the words they negate. [məiʌ] is used to negate the existential verb [jəuʌ] and verbs in the perfective aspect, whereas [puʌ] is used with other verbs and adjectives:

- (15) a. wəʌ puʌ k^hyʌ
 1SG NEG go
 'I am not going.'
- b. wəʌ məiʌ jəuʌ k^hyʌ
 1SG NEG EXIST go
 'I haven't gone.'

[puʌ] changes to the second (rising) tone before another fourth-tone (falling) word. This change is not reflected in the orthographies.

The perfective particle [lə] is obligatorily omitted when the verb is negated by [məiʌ]. Optionally, [jəuʌ] 'exist' may be inserted between the negative [məiʌ] and the main verb in the perfective (15b). In addition, it should be noted that [puʌ] is also used to negate verbs that are followed by the [lə] that is inchoative.

4.10.3.8 Word order and grammatical relations

The basic word order is SVO as in English; modifiers, including relative clauses, generally precede the words they modify. Since direct objects are not distinguished from subjects or topics by case markers, word order is important and relatively fixed:

- (16) a. wəʌ maʌ tʃaŋʌ sanʌ
 1SG scold Zhāng Sān
 'I scold Zhāng Sān.'
- b. tʃaŋʌ sanʌ maʌ wəʌ
 Zhāng Sān scold 1SG
 'Zhāng Sān scolds me.'

Grammatical relations such as locative and instrumental are usually signaled by means of preposed verbs instead of case suffixes (as in Russian and Finnish) or by prepositions (as in English). Such a use of a series of verbs, none of them syntactically subordinated to each other, is called a SERIAL VERB CONSTRUCTION.

- (17) a. wəʌ tswəʌ fəiʌkiʌ tauʌ pəiʌkiŋʌ k^hyʌ
 1SG sit airplane arrive Běijīng go
 'I am going to Peking by plane.'

- b. wəʌ juŋʌ jauʌʃi k^haiʌ mənʌ
 1SG use key open door
 ‘I open the door with a key.’
- c. wəʌ paʌ nəiʌ pənʌ ʃuʌ suŋʌ kəiʌ t^haʌ lə
 1SG take that CLF book send give 1SG PFV
 ‘I sent that book to him.’

In (17c) [paʌ] acts as an indicator of the direct object when it appears before the verb. This construction is used primarily when the direct object is a definite noun phrase.

4.10.3.9 Passive voice

The passive voice, too, is signaled by a special verb and has ADVERSATIVE value, that is, it implies that the PATIENT – the person who undergoes the action expressed in the main verb – was in some way unpleasantly affected by it. The AGENT of the passive construction – the person who is doing the action – is usually marked by the verb [pəiʌ], literally ‘to suffer’:

- (18) wəʌ pəiʌ t^haʌ-mən k^hanʌkʃənʌ lə
 1SG suffer 3-PL see PFV
 ‘I was seen by them (and that was unpleasant for me).’

Note that even intransitive verbs can be passivized in Chinese:

- (19) wəʌ pəiʌ t^haʌ p^hauʌ lə
 1SG suffer 3SG run_away PFV
 ‘He ran away on me.’

The adversative passive, like the noun classifiers discussed earlier, is an areal feature found in a continuum that stretches from Southeast Asia through China and on to Japan and Korea.

4.10.3.10 Interrogative

Interrogative sentences that contain question words (i.e. content questions) do not add anything or change their word order (20a). To change statements into polar questions, one may either add an interrogative particle such as [ma] at the end of the sentence (20b) or use a verb_i-not-verb_i construction (20c):

- (20) a. ʃəiʌ k^hyʌ
 who go
 ‘Who is going?’

- b. niʌ kʰyʌ ma
 2SG go Q
 ‘Are you going?’
- c. niʌ kʰyʌ puʌ kʰyʌ
 2SG go NEG go
 ‘Are you going?’

4.10.3.11 Relative and other subordinate clauses

Relative clauses precede the head nouns and are connected to them by the subordinating particle [tə], which in other contexts serves as a possessive marker:

- (21) tswəʌtʰjənʌ laiʌ tə ɹənʌ ʃiʌ wəʌ tə pʰəŋʌjəu
 yesterday come SUBORD person be(COP) 1SG POSS friend
 ‘The person who came yesterday is my friend.’

Conditional clauses are usually introduced by [jauʌʃi] ‘if’, but this subordinating conjunction can optionally be omitted:

- (22) jauʌʃi niʌ puʌ jauʌ kʰyʌ wəʌ jəʌ puʌ jauʌ kʰyʌ
 if 2SG NEG want go 1SG also NEG want go
 ‘If you don’t want to go, I don’t want to go.’

Some temporal subordinate clauses are syntactically relative clauses with [ʃiʌxəu] ‘time’ as the head noun:

- (23) tʰaʌ laiʌ tə ʃiʌxəu niʌ xaiʌ məiʌ tauʌ ʃaŋʌxaiʌ
 3SG come SUBORD time 2SG still NEG arrive Shànghǎi
 ‘When he came you still had not arrived in Shanghai.’

Other temporal clauses consist of a subordinate clause unmarked by [tə] but followed by a time adverbial, which is used as a kind of postposition:

- (24) tʰaʌ laiʌ-lə iʌxəuʌ wəʌ-mən kʰyʌ kʰanʌ tʃənʌiŋʌ
 3SG come-PFV afterwards 1-PL go look movie
 ‘After he comes we will go see a movie.’

4.10.4 SAMPLE TEXTS

The first selection is a Táng dynasty poem written by 張繼 Zhāng Jí, who lived in the eighth century AD. It illustrates the terse beauty of Chinese poetry along with an interesting typological point: Chinese classical poetry is the best example of

analytical morphology. Almost all grammatical relations between words are signaled primarily by word order, not by affixes or even particles. Particles existed in Chinese, but were avoided in Chinese classical poetry. Because economy of expression was highly esteemed, particles in a poem would have been considered clutter. Of course, one must keep in mind that the language of poetry may differ significantly from the spoken colloquial or even literary prose.

The top line of these transcriptions are in traditional Chinese characters. The pronunciation given is that of modern standard Mandarin. The actual language spoken by Zhāng Jí was Middle Chinese, which had quite a different pronunciation. To give some idea of how much the pronunciation has changed since the Táng dynasty, in Late Middle Chinese the title of this poem was pronounced [fjyŋ kfiaw jiaʋ pfiak] (Pulleyblank 1991).

Each line of the poem is seven syllables long. The rhyme scheme is *aaba*, although in modern pronunciation [tʃʰwanʔ] ‘boat’ does not rhyme very well with [tʰjənʔ] ‘sky’ and [mjənʔ] ‘sleep’; in Middle Chinese, all these words ended in [an]. There are also complicated rules governing the sequences of tones in each line. Our interlinear free translations are adapted from Chao (1957: 274), who cites C. W. Lu’s *On Chinese poetry*, published in Beijing in 1935.

- (25) 楓橋夜泊 Fēng qiáo yè bó

fəŋʅ kʰjauʔ jəʋ pɥəʔ
maple bridge night moor

‘Night Mooring at Maple Bridge’

- (26) 月落烏啼霜滿天。Yuè luò wū tí shuāng mǎn tiān.

ɥəʋ lwəʋ uʅ tʰiʅ [ɥwaŋʅ] manʅ tʰjənʅ
moon fall raven cry frost fill sky

‘The moon goes down, a raven cries, frost fills the sky.’

- (27) 江楓漁火對愁眠。Jiāng fēng yú huǒ duì chóu mián.

kjaŋʅ fəŋʅ yʅ xwəʅ twəiʅ tʃʰəuʅ mjənʅ
river maple fish fire face sad sleep

‘River maples, fishing fires, facing troubled sleep.’

- (28) 姑蘇城外寒山寺 Gūsū chéng wài Hán shān sì

kuʅsuʅ tʃʰəŋʅ waiʅ xanʅ ʃanʅ siʅ
Gūsū City outside Cold Mountain temple

‘Outside Gūsū city, Cold Mountain Temple’s’

[kuʅsuʅ] is an ancient name for present-day Sūzhoū, a town near Shànghǎi.

- (29) 夜半鐘聲到客船。yè bàn zhōng shēng dào kè chuán.

jəʋ panʅ tʃʰuŋʅ ʃəŋʅ tauʅ kʰəʋ tʃʰwanʅ
night half bell sound reach guest boat

‘midnight bell’s sound reaches the visiting boats.’



FIGURE 4.4 In 1981 the People's Republic of China issued a set of five stamps depicting the story of the foolish man from Chu who lost his sword. The first stamp in the set includes the text of this story in the original Classical Chinese. Designer: 潘可明 Pān Kěmíng.

The Cold Mountain Temple and the Maple Bridge next to it are still in existence and have become quite an attraction, especially for Japanese tourists who come to visit this famous temple, hear its bell, and buy hanging scrolls on which this poem, well known in Japan, is beautifully inscribed by local calligraphers. Japanese people study Chinese classical poetry just as in the old days students in the West used to study Latin poetry. On New Year's Eve, thousands of Japanese tourists jam the street near the temple to hear the midnight bell as it tolls in the New Year.

The second selection is a short prose story written in modern literary Mandarin. It is actually a modern retelling of a story from one of the Chinese classics (Liu 1960: 15); the original text can be seen in the first stamp in Figure 4.4. The first lines of our glosses give the Chinese text in the current, official, simplified characters.

- (30) 有一个楚国人过江的时候，他的剑从船上掉到水里去。

Yǒu yī ge Chǔguó rén guò jiāng de shíhòu, tā de jiàn cóng chuán shàng diào dào shuǐ lǐ qù.

jəu˥ i˧ kə ʈʰu˥u˥-kwə˧˥-ʈən˥ kwə˧ kjaŋ˥ tə ʃi˧xəu tʰa˧ tə kjaŋ˥
EXIST one CLF Chǔ-country-person cross river SUBORD time 3SG POSS sword

ʈʰu˥uŋ˥ ʈʰwan˥ ʃaŋ˥ tʃau˥ tau˥ ʃwəi˥ li˥ kʰy˥
from boat on(POST) fall arrive water in(POST) go

‘As a certain man from the country of Chǔ was crossing a river, his sword fell into the water from the boat.’

POST = postposition.

[ʈʰu˥u˥]. The name of an ancient Chinese kingdom located in Central China.

[kjaŋ˥]. In Classical Chinese literature [kjaŋ˥] referred specifically to the Yangtze River. In modern Chinese it may refer to any river.

[kʰy˥] ‘go’ here has the function of signaling movement away from the speaker, in much the same way as the German verb prefix *hin-*.

- (31) 他马上在船边上刻了一个印子，说：

Tā mǎshàng zài chuán biān shàng kè le yī ge yinzi, shuō:

tʰa˧ ma˧ʃaŋ˥ ʈsai˥ ʈʰwan˥ pʃən˥ ʃaŋ˥ kʰə˧˥-lə i˧ kə in˥ʈsi ʃwə˧
3SG immediately be_located boat side on(POST) carve-PFV one CLF mark say

‘He immediately carved a mark at the side of the boat, saying.’

[ma˧ʃaŋ˥]. Literally, ‘on a horse’.

- (32) 我的剑是从这儿掉下去的。

“Wǒ de jiàn shì cóng zhèr diào xiàqù de.”

wəʌ tə kʲənʌ ʃiʌ ʈʰuŋʌ ʈʰaɪʌ tʃauʌ xjaʌ-kʰyʌ tə
1SG POSS sword COP from here fall down-go SUBORD

‘“My sword fell from here.”’

The head noun here is understood to be the same as the topic ‘sword’. Therefore, literally the sentence is ‘My sword is (the one) that fell down from here’. This type of construction is often used to express past actions.

- (33) 船停了，他就按照他刻了印子的地方下水去找。

Chuán tíng le, tā jiù ànzhào tā kè le yinzi de dìfāng xià shuǐ qù zhǎo.

ʈʰwanʌ ʈʰiŋʌ lə ʈʰaʌ kʲəuʌ anʌʈʰauʌ ʈʰaʌ kʰəʌ-lə inʌʈʰsi tə tiʌʃaŋʌ
boat stop PFV 3SG then based_on 3SG carve-PFV mark SUBORD place

xjaʌ ʃwəiʌ kʰyʌ ʈʰauʌ
descend water go seek

‘After the boat stopped, he went down into the water to seek his sword based on the position of the mark he had carved.’

[kʲəuʌ] ‘then’ is a sequential action marker.

- (34) 船已经动了，可是剑没动。

Chuán yǐjīng dòng le, kěshì jiàn méi dòng.

ʈʰwanʌ iʌkiŋʌ tuŋʌ lə kʰəʌʃi kʲənʌ məiʌ tuŋʌ
boat already move PFV but sword NEG.PFV move

‘The boat had already moved, whereas the sword had not.’

- (35) 找剑像这样找法不是糊涂吗？

Zhǎo jiàn xiàng zhè yàng zhǎo fǎ bù shì hútu ma?

ʈʰauʌ kʲənʌ xjaŋʌ ʈʰəʌ jaŋʌ ʈʰauʌ faʌ puʌ ʃiʌ xuʌʈʰuʌ ma
seek sword like this kind seek method not COP silly Q

‘Isn’t this a muddle-headed way of looking for the sword?’

4.11 Sketch of Classical Tibetan

4.11.1 GENETIC RELATIONSHIP AND GENERAL BACKGROUND

Tibetan is a member of the Himalayish subbranch of the Tibeto-Burman branch of Sino-Tibetan. It is spoken by over six million speakers in a variety of dialects, the chief of which is the dialect of Lhasa. The large majority of the speakers live in the Tibet Autonomous Region of China; the rest are scattered in Chinese provinces adjoining Tibet, especially Sichuān and Qīnghǎi, as well as in northeast India, Nepal, and Bhutan.

The earliest written records in Tibetan date from the seventh century AD. It was about that time that Buddhism was imported into Tibet from India along

with the writing system. For many centuries thereafter, the culture and language of Tibet were strongly influenced by Indian culture. Although relatively few words from Sanskrit and other Indic languages were borrowed into Tibetan directly, there are many CALQUES from these languages, especially Sanskrit. A calque is a loan translation: a word borrowed from another language by translating its morphological components into morphemes of the borrowing language. An example of a typical calque is the following epithet of Buddha, which was interpreted as meaning ‘one who has gone like this’:

- (36) a. Sanskrit: Tathāgata
 tet^ha:-gə-t-ə
 thus-go-PST.PTCP-M
- b. Tibetan:
 de-bzin-gcegs-pa
 that-like-went-NOUN

In turn, Tibetan has strongly influenced Mongolian ever since Mongols embraced Tibetan Buddhism as their religion.

There is an extensive literature that consists mostly of Buddhist texts and commentaries. Because some of the Buddhist texts were preserved only in their Tibetan translation and because Tibetan Buddhism (Lamaism, Tantric Buddhism) has developed in a unique fashion, the Tibetan language is extensively studied by Buddhologists. Purely secular literature has been almost nonexistent until recently except for some historical chronicles, folk tales, and poetry.

Currently the Tibetan language in areas under Chinese control is undergoing rapid changes. Thousands of new terms are being introduced into the language in connection with the importation of new political and scientific concepts from Chinese-speaking China. Some of these loanwords are direct loans from standard Chinese (Mandarin), but the majority of them are also calques from that language.

The modern standard spoken language is based on the speech of Lhasa, the capital of the Tibet Autonomous Region. Although there are many dialects of Tibetan, some of which are probably not mutually intelligible, the written language is fairly uniform throughout Tibet. This is because the written language does not reflect any modern dialect but is essentially a modified version of the classical language of the seventh century AD. The modification involves infusion of more modern vocabulary and a somewhat confused attempt to reflect the verb morphology of modern Lhasa Tibetan. This amalgam of ancient and modern features is called *Modern Written Tibetan*. In this sketch only the more conservative Classical Tibetan rather than Modern Written Tibetan or the colloquial dialect of Lhasa will be described.

In spite of the very large divergences between the written language and modern VERNACULARS (local spoken dialects), which impede universal literacy in Tibet, there are no plans for major reforms of the written language in the near future. Chinese linguists consider that no reforms are possible until the dialect

of Lhasa is more widely spoken and understood, since the written language of the future will have to be based on this dialect. To switch to Lhasa colloquial as the written language now would mean that large numbers of Tibetans would be unable to understand this new written language, and to allow the number of different written languages based on different colloquial standards would hinder communication among different regions.

Table 4.14 (mapped on Figure 4.2) lists the major Tibetan language varieties as described by Tournadre (2005). Traditionally, these have been described as dialects of Tibetan, but the more recent trend is to describe most of these varieties as separate languages, on the grounds of lack of mutual intelligibility between most of them. Ethnologue provides separate language codes for 39 forms of Tibetan.

The cladogenesis of the Tibetan dialects or languages has not yet been worked out. But one useful approach to grouping the dialects is in terms of how conservative or innovative they are in their phonology. The most conservative languages are the Ladakhi-Balti dialect group. These are spoken in the region historically called *Ladakh*, which is just to the west of Tibet; in modern terms, the region encompasses Baltistan in Pakistan and Jammu and Kashmir in India. The phonology of the Ladakhi-Balti languages, especially Balti, is close to that of Classical Tibetan.

The next most conservative set of dialects are the Amdo group. These are spoken mainly in Tibetan autonomous prefectures in Qīnghǎi province, just northeast

TABLE 4.14

Tibetan languages

Name	Size	Location
Central Tibetan (BOD)	7	Tibet
· Ü		Lhasa
· Tsang		Shigatse
· Loke (LOY)		Nepal
Amdo (ADX)	7	east Tibet
Kham-Hor		east
· Khams (KHG)	7	Tibet
Ladakhi-Balti		west
· Balti (BFT)	6	Pakistan
· Ladakhi (LBJ)	6	India
· Zangskari (ZAU)	5	India
· Purik (PRX)	5	India
Lahuli-Spiti		west
· Spiti Bhoti (SPT)	5	India
Sherpa-Jirel		southwest
· Sherpa (XSR)	6	Nepal
Kyirong-Kagate		southwest
· Kagate (SYW)	4	Nepal
Dzongkha-Lhokä		south
· Dzongkha (DZO)	6	Bhutan

of Tibet itself. Amdo dialects have lost [s] at the end of syllables, but are otherwise fairly conservative.

The other dialect groups listed are much more innovative. Not only do they drop final [s], but they also drop all of the consonants that in Classical Tibetan could be prefixed before the root. For example, the classical word for 'door', [sgo], has changed minimally, to [zgo] in the Ladakhi-Balti languages and to [rgo] in Amdo, but in the innovating dialects, the original [s] is gone entirely. These same, innovating, dialects have also undergone TONOGENESIS, the birth of phonological tone systems. After a voiceless consonant or certain clusters, syllables developed a high tone; otherwise, they developed a low tone. In many dialects, this tone became phonemically significant after a later change that devoiced the consonants. For example, in the innovating Ü, classical [bka] became [ka^h] and [sga] became [ka^l].

The conservative dialects lie far from the Tibetan capital; most speakers of Amdo and Ladakhi-Balti languages do not even reside within Tibet. The dialects of Central Tibet have undergone many more radical sound changes, especially involving consonants, than the dialects in the peripheral regions. Central Tibetan includes the Ü (Dbus) and the Tsang (Gtsang) subgroups. The former includes the dialect of Lhasa, the capital of Tibet, and the latter includes the dialect of Shigatse, the second largest city. The dialect of Shigatse is more conservative and is closer to the written language than that of Lhasa. The Kham-Hor dialects to the north and east have also innovated tone, and are generally treated in China as a separate language from Central Tibetan.

The Dzongkha-Lhokä dialects extend south into India (Sikkim) and Bhutan. The characteristic feature of these dialects is a vowel apocope that reduces disyllabic lexical items to monosyllables: original [sdepa] 'regional official' is pronounced [dep] in these dialects. In the southwest, Sherpa is spoken in Nepal by the Tibetan people famous for their expert mountaineering guides. Toward the west into India, the Lahuli-Spiti group of dialects are considered innovating dialects, in that they have developed tone and lost final [s], but they have preserved other final consonants, namely [l] and [d], more than most other innovating dialects.

4.11.2 ORTHOGRAPHY

The Tibetan script is an adaptation of an Indic script used in northern India in the seventh century AD. It still resembles such Indic scripts as the Devanagari, which is used for Hindi and other languages of north India and Nepal. In turn, the Tibetan script was the basis for the 'phags-pa alphabet introduced to China during the Mongolian Yuán dynasty (1271–1368 AD). It was designed by a Tibetan lama to serve as a universal alphabet for all the different languages of the Mongol empire spoken at the time, including Chinese. However, after the expulsion of the Mongol rulers from China, this alphabet also fell into disuse. Before its total demise, however, it may have influenced the creation of hangeul, the native Korean alphabet (§4.1.4).

The Tibetan writing system is phonemographic. Like Indic scripts and other scripts based on them, it is an abugida (§2.1.3.4). All vowels but [a] are written with a distinct, explicit diacritic. If a syllable is written with no vowel symbol, the implicit vowel [a] is assumed. As we shall see shortly, syllable boundaries are explicit in Tibetan, but knowing exactly where to put the [a] can still be tricky, because the Tibetan syllable allows consonant clusters in both the onset and the coda.

Example (37) gives the Tibetan letters along with their pronunciation in IPA (Jacques 2012). They are arranged in Tibetan alphabetic order. Note how this order, reflecting Indic linguistic tradition, shows excellent understanding of articulatory phonetics. This is perhaps clearest in the top half of the list (37a–d). The consonants are arranged according to their place of articulation from the back part of the mouth to the front. Within each series they are arranged according to the manner of their articulation: first the voiceless sounds, in the order unaspirated then aspirated, followed by the voiced sounds, first oral, then nasal.

- (37) a. ཀ ཁ ག ཅ
 k k^h g ŋ
 b. བ ཏ ཐ ཌ ཎ
 ṽ ṽ^h ḍṽ ṇ
 c. ཉ ཏ ཏ ཏ
 t t^h d n
 d. པ པ པ པ
 p p^h b m
 e. ཚ ཛ ཛ ཛ ཛ ཛ ཛ
 ṽ ṽ^h ḍṽ w ṣ z ṽ
 f. ལ ལ ལ
 j r l
 g. ཤ ཤ ཤ ཤ
 ç s h ?

Row (37e) is an interpolation of letters that were invented to represent sounds not found in Sanskrit. The history of the letter ཛ has long been debated, but in Classical Tibetan it appears to often function as a graphic MARKER in the sense of Venezky (1999): something that has the shape of a letter but the function of a diacritic. The function of this marker is varied, as we will address below. Its most substantial function from a phonological point of view is to indicate that the following consonant is prenasalized. These consonants begin as nasals but end as orals at the same place of articulation. There is no official IPA symbol for prenasalization, but we will follow here a common convention of placing the symbol [ṽ] before the symbol that represents the oral consonant. The place of articulation of a prenasalized consonant is indicated by the base letter; thus ཛག [ṽg] is velar all the way through.

As already pointed out, the vowel [a] is unmarked in Tibetan script. Other vowels are written by diacritics that are added to the letter whose sound precedes the vowel. All the vowels of Tibetan are illustrated in (38) with the consonant [m]. Note that the vowel diacritic is always read off after the consonant, regardless of whether the diacritic appears on top or underneath:

- (38) མ་ མི་ མུ་ མེ་ མོ་
ma mi mu me mo

The raised dot marks the syllable boundary and is very important for determining where the [a] should be understood. For example, in the following combinations, the presence of the syllable boundary symbol indicates whether we are dealing with a single syllable, in which case only one [a] has to be supplied in reading (39a), or with two syllables, in which case two [a] have to be read (39b):

- (39) a. ལས་ [las]
b. ལ་ས་ [la.sa]

In this book, our transcriptions will normally supply the IPA syllable-divider “” only when syllable boundaries are potentially ambiguous.

One way to indicate the position of [a] in the syllable is by stacking letters vertically. Such stacks are read from top to bottom, and no vowels are ever read between the component consonants:

- (40) a. གལ་ [gal]
b. གལ་ [gla]
c. གལི་ [gli]

Stacked consonants mostly look like miniature versions of what the consonant looks like by itself, but a few consonants take special forms when they appear above (41a) or below (41b–41c) another letter, here illustrated with ག [g].

- (41) a. གོ་ [rga]
b. གྲ་ [gra]
c. གླ་ [gwa]

There is also a special form that can be written under some consonants to indicate palatalization (42a). Traditionally this is considered to be just a variant shape of the consonant ག, but note the crucial contrast between the palatalized [gʲ] (42a) and the sequence [gʲ] (42b). We know that this is not just an arbitrary peculiarity of the writing system, because in the modern dialects these forms have different pronunciations: in Lhasa, the palatal stop [c] for གླ་ versus [j] for གལ་.

- (42) a. གྲ [gʰa]
 b. གཤ [gja]

Most consonants are written linearly (horizontally) without stacking, even if they form an onset cluster. In some cases, that could, in principle, look the same as the spelling for a sequence consisting of an onset consonant, implied [a], and coda consonant. To forestall that latter interpretation in cases of potential ambiguity, the marker འ is used:

- (43) a. བག [bag]
 b. བགའ [bga]

The same marker is used as the basis of the second vowel of a diphthong:

- (44) a. འ [ʔa]
 b. འཤ [ʔai]

In addition to the dot marking the syllable boundary, Tibetan orthography uses a number of punctuation marks, the chief of which is a vertical line, which marks a pause. Word boundaries are not noted at all.

4.11.3 PHONETICS AND PHONOLOGY

Classical Tibetan appears to have had essentially the same phoneme inventory as is listed in (37) and (38). There are no diphthongs within morphemes; all the diphthongs are derivable from underlying sequences of monophthongal morphemes. There is no provision in the orthography for marking tones, and it is believed that Classical Tibetan was not a tone language.

The syllable structure of Classical Tibetan allows a large number of consonants to begin a syllable. For example, [bsgr] is a possible syllable-initial cluster. The structure of the rest of the syllable is less complex: a syllable may end in a vowel, or a vowel followed by a nasal, a voiced stop, a liquid, or [s]. Voiced stops after a vowel may be followed by [-s] to yield the final consonant clusters [bs] and [gs].

4.11.4 MORPHOLOGY AND SYNTAX

By the time of the earliest inscriptions in Tibetan, much of the earlier affixal morphology had become fossilized and opaque. For example, it is possible to isolate a prefix [g-] in such Tibetan numbers as [gciq] 'one', [gɲis] 'two', and [gsum] 'three', but unless one resorts to comparative evidence, it is not possible to assign any specific meaning or function to this prefix. What will be described in this section is only the more productive and functional morphology of Classical Tibetan.

4.11.4.1 Noun phrase

A noun phrase headed by a full noun (not a pronoun) has one of the following structures:

- (45) a. noun (adjective) (demonstrative) (plural) case
 b. noun (adjective) (plural) (indefinite) case

The parentheses indicate that only the head noun and the case particles are obligatory constituents; everything else is optional. Unlike in Russian, adjectives do not normally change their form to agree with the gender of the noun they modify. An example of pattern (45a):

- (46) mi t̪ɕ^henpo de rnam=la
 man large DEM PL=DAT
 'to those large men'

For the most part, Tibetan adjectives function grammatically like nouns and can themselves serve as heads of noun phrases.

4.11.4.2 Number

The most common plural marker is [rnam]. Other plural markers are [dag], [t̪ɕag], and a few other rarer ones. The first, [dag], is very often used to mark plural when directly addressing people and, in texts translated from Sanskrit, represents Sanskrit dual number, which is not a morphological category in Tibetan itself.

The singular-plural distinction is strictly observed only in the pronouns. With nouns, plural is only optionally marked, and is usually omitted for inanimate nouns.

4.11.4.3 Indefinite particle

There is no definite article in Tibetan, although a sense of definiteness is often signaled by the demonstrative [de] 'that'. Indefiniteness is signaled by the particle [t̪ɕig], which is related to the numeral [gt̪ɕig] 'one'. Unlike [gt̪ɕig], which is invariable, the indefinite particle changes its initial consonant depending on the final segment of the preceding morpheme:

- (47) a. [zig] after SONORANTS (nasals, liquids, vowels)
 b. [ɕig] after [s]
 c. [t̪ɕig] after [b], [d], [g]

After a plural marker or a number, the indefinite particle can be translated as 'about' or 'some':

- (48) dge.sloŋ l̪ja br̪g¹a zig
 monk five hundred INDF
 'about 500 monks'

4.11.4.4 Cases

Traditionally, the term *case* has applied to how noun inflections correspond to grammatical relations, as in Russian and Finnish. In Tibetan, there are no noun inflections. Instead, the same type of syntactic information is expressed by particles at the end of the noun phrase. These particles can readily be identified as words rather than inflections because they follow whatever word is at the end of the phrase, whereas inflections attach to specific word classes, like nouns. For want of a better term, linguists tend to refer to these particles as *case particles*, emphasizing the syntactic function of case over its morphological characteristics. Tibetan case particles do, however, have some characteristics in common with case suffixes: they attach to the end of a word and are subject to morphophonemic changes, as shown in Table 4.15. Particles that behave phonologically like suffixes are called ENCLITICS. In linguistic glosses, enclitics are set off by double hyphens “=” instead of single hyphens “-”.

ABSOLUTIVE

Like several other languages we have encountered, including Basque and Georgian, Tibetan is an ergative language. No overt marker is used for the subject of an intransitive verb and for the direct object of a transitive verb.

GENITIVE

Nouns are usually the first element in a noun phrase. But dependents may instead precede the noun, in which event they are separated from the head noun by a genitive marker. Relative clauses regularly take this prenominal position (49), and

TABLE 4.15

Tibetan grammatical relation markers

	Case	When after	Form
ABS	absolutive	any	∅
GEN	genitive	vowel	i
		velar (g, ŋ)	gi
		other sonorant (l, r, m, n)	gʲi
		else (d, b, s)	kʲi
ERG	ergative	vowel	s
		velar	gis
		other sonorant	gʲis
		else	kʲis
INE	inessive	any	na
DAT	dativ	any	la
ABL	ablative	any	las
ELA	elative	any	nas
ALL	allative	vowel	r, ru
		s	su
		other sonorant (l, r, m, n, ŋ), d	du
		else (g, b)	tu

adjectives and demonstratives may optionally be fronted to this position as well. A prenominal noun dependent is typically a possessor.

- (49) dŋul btaŋ-ba=i mi
 silver gave-NOUN=GEN man
 ‘the man who gave the money’

ERGATIVE

This case marks the subject of a transitive verb. It also marks the instrument that an action is performed with, like the Russian instrumental case. From the word order and the context it is usually clear which of the noun phrases marked with the ergative is the subject and which is the instrument:

- (50) rgʲal.po=s mi de ral.gri=s bsad
 king=ERG man DEM sword=ERG killed
 ‘The king killed that man with a sword.’

When added to nominalized clauses, this case marker usually indicates cause:

- (51) ɕintu bka.ba jin-pa=s
 very difficult be-NOUN=ERG
 ‘because (it) is very difficult’

The marker for the ergative case is a combination of the genitive marker and [-s]. Two other case markers share this same property of being like another marker, but ending in [-s]. What these all have in common appears to be the concept of SOURCE: where something comes from.

DATIVE

The dative case marks the indirect object of a verb, often a recipient. It can also mark location at or toward a place. With some verbs it marks the sole object:

- (52) a.ma=la ltos
 mother=DAT look(IMP)
 ‘Look at mother!’

One of its other uses is to mark the possessor in constructions like the following:

- (53) ŋa=la ralgri jod
 1SG=DAT sword EXIST
 ‘I have a sword.’
 literally, ‘To me there is a sword.’

– a construction familiar from Russian and Finnish.

LOCATIVE CASES

Like Finnish, Tibetan has several cases that primarily mark location, in addition to the dative. All of them have a core meaning, which is often overwhelmed by many other secondary meanings. The inessive case basically marks a stationary position in time or space, especially one in an enclosed area.

- (54) a. $de=na$
 DEM=INE
 ‘then’ or ‘there’
- b. $de=i$ $\widehat{ts}^h e=na$
 DEM=GEN time=INE
 ‘at that time’

The allative case has a large variety of functions, some of which duplicate those of other cases:

- ✗ to mark motion toward or into something (like the dative)
- ✗ to indicate location in space and duration in time (like the inessive)
- ✗ to make manner adverbs out of adjectives
- ✗ to mark what something becomes or turns into
- ✗ to indicate purpose

and a number of other uses that cannot be neatly summarized.

- (55) a. $gronk^j h er=du$ $^n d\check{z}ugpa$
 city=ALL enter
 ‘enter the city’
- b. $k^j h im=du$ jod
 house=ALL EXIST
 ‘(He/she) is at home.’
- c. $m^j ur=du$
 quick=ALL
 ‘quickly, soon’
- d. $\widehat{t\check{c}a\eta}$ $dug=tu$ $g^j ur$
 beer poison=ALL became
 ‘Beer became poison.’
- e. $t^h os.pa=r$ $s\eta anpa$
 hear=ALL pleasant
 ‘pleasant to hear’

- f. *ṅinpa=r*
 day=ALL
 ‘during the day’

Tibetan has two cases that mark source of motion. The ablative and the elative were originally distinguished by the fact that the latter suggests that the moving object came from inside something else, although the distinction was rarely observed very strictly in Classical Tibetan.

- (56) *rta=nas ṅuṅ*
 horse=ELA fell
 ‘He fell off a horse.’
 [las] is also acceptable.

The ablative marker [las] is also used in comparative constructions to mark the standard to which something is being compared:

- (57) *gser=las rdo.rḍze dkonpa jin*
 gold=ABL diamond precious COP
 ‘Diamonds are more precious than gold.’

Note that adjectives have no special inflection to signal the comparative degree.

As cases that express source, the use of the ablative and elative may overlap with the instrumental use of the ergative:

- (58) *lagpa nas mi de ṅḍzinpa*
 hand ELA man DEM grasp
 ‘She grasped that man by the hand.’
 (Literally, ‘took that man by means of the hand.’)

Note that the hand in question does not belong to the subject of the sentence. Those who are familiar with Japanese will note that uses of [las] and [nas] parallel the uses of Japanese [jori] and [kara] respectively.

4.11.4.5 Verb phrase

Tibetan verbs are not inflected for person or number. There are four basic forms in the Classical Tibetan verb paradigm, as illustrated here with the stem [sgrim]

'hold fast'. This chart shows, for example, that [sgrim] with no affixes is present tense and [bsgrim] is future tense (FUT).

(59)	Tense/Mood	Prefix	Stem	Suffix
	PRS		sgrim	
	PST	b-	sgrim	-s
	FUT	b-	sgrim	
	IMP		sgrim	-s

The [b-] prefix that appears in the past and future forms is generally not affixed to verbs that do not have any derivational prefixes. Moreover, this prefix as well as the suffix [-s] are frequently deleted by PHONOTACTIC rules: phonological rules that prevent certain sequences of phones. For example, there is a rule preventing two labials from occurring in the same onset cluster, and another rule preventing [b] from appearing before a prefix consonant that is a stop. This deletion of affixes often leads to homophonous forms within the same paradigm. Ambiguous forms are often replaced by constructions based on the present tense form in combination with various auxiliary verbs.

A fairly large number of verbs exhibit unpredictable vowel apophony:

- (60) 'do':
- a. PRS [b'ed]
 - b. PST [b'as]
 - c. FUT [b'a]
 - d. IMP [b'os]

The verb [b'ed] 'do' is also irregular in that it has [-d] in the present. Verbs whose present tense form contains either [i] or [u] very rarely exhibit vowel apophony. For other verbs, among the most common apophonies are changing the vowel to [a] in the past and future and changing it to [o] in the imperative.

For the sake of brevity, numerous exceptions and some very common affixal and apophonic patterns have been omitted from this description, but the foregoing summary holds true for a very large number of Tibetan verbs.

An import derivational prefix is [s-], which turns stems into transitive verbs. These transitive verbs are often CAUSATIVE – their subjects cause a change of state or cause somebody to do something:

- (61) a. [bugpa] 'hole' → [sbugpa] 'to pierce'
 b. [ʔg'urpa] 'to become' → [sg'urpa] 'to transform something, to cause to become'

4.11.4.6 Subordinate clauses

There are several enclitic particles that signal that a verb is in a subordinate clause. Two very general markers, which we will gloss as =SUBORD, are [ste] and [t̥ɕiŋ]. They attach to finite verbs, most commonly past and present tense forms. [t̥ɕiŋ] exhibits the same morphophonemic changes as that of the indefinite particle (§4.11.4.3). [ste] has three allomorphs:

- (62) a. [de] after [d]
 b. [te] after other dental consonants
 c. [ste] elsewhere

Their use is similar to that of participle clauses in English. Their actual meaning is very general; usually a good English translation requires one to infer the semantics from context. Often a good translation is to supply a word like *after*, *because*, *as*, or *when* before the subordinated clause, or to follow it with a word like *then* or *and*.

- (63) a. de=s rdo ʒig blaŋs=te ʰpʰaŋs.pa
 DEM=ERG stone INDF took=SUBORD threw
 ‘He took a stone and threw it.’
 Literally: ‘He, having taken a stone, threw.’
- b. rgʲal.po=s kʰrims b̥t̥ɕa=ste bzaŋ=la bʲa.dga bster
 king=ERG law enacted=SUBORD good=DAT reward bestowed
 ‘Because the king enacted the law, the good people were rewarded.’
 Literally: ‘The king having enacted the law, the good people were rewarded.’
- c. de=i druŋ na tʰa.gapa ʒig tʰags.ʰtʰag t̥ɕiŋ ʰdug.pa de=i
 DEM=GEN front INE weaver INDF weave SUBORD sit-NOUN DEM=GEN
 steŋ=du ʒuŋ
 top=ALL fell
 ‘(He) fell on top of a weaver, that (one) who was sitting and weaving in front of it (the wall).’

Many of the noun case enclitics can be used in a different capacity after verbs. In such a position, [na] marks a conditional clause (‘if’). After present and past forms of the verb, either [las] or, more commonly, [nas] is used to mark subordination of that verb’s clause to another verb in the sentence. This subordination usually expresses time, but sometimes cause.

- (64) ba.glaŋ sgo gʒan=du soŋ nas stor ro
 bull door other=ALL went ELA became_lost FP
 ‘After the bull went through another door, he became lost.’
 [ro] is a sentence final particle (§4.11.4.7).

4.11.4.7 Main verb markers

INTERROGATIVE PARTICLE

This marker has two allophones:

- (65) a. [am] after vowels
 b. C[am] after C; i.e. copy of the preceding consonant + [am]

It marks interrogation in polar questions. Content questions have the normal word order and no special marking.

- (66) k^hod=kⁱs glaŋ br̥nas=sam
 2SG=ERG bull borrowed=Q
 ‘Did you borrow the bull?’

SENTENCE-FINAL PARTICLE

- (67) a. [o] after vowels
 b. C[o] after C (i.e. copy of the preceding consonant + [o])

The final particle (FP) marks an emphatic end of a sentence. Because the main verb phrase comes at the end of the Tibetan sentence, the particle can be considered to be the main verb phrase marker or even the main verb marker. Sometimes it has the force of a copula verb when the latter has been optionally deleted and there are no other sentence-final verbs; some scholars believe that it was originally a copula verb itself:

- (68) a. ŋa=s glaŋ br̥nas=s
 1SG=ERG bull borrowed=FP
 ‘I borrowed the bull.’
 b. de=i gdoŋpa mdz̥es kjaŋ=ŋo
 DEM=GEN face beautiful even=FP
 ‘His face is even beautiful.’

4.11.4.8 Existential, locational, and copula verbs

Existential, locational, and copula verbs are not inflected for tense and do not have imperative forms. The same is true of many other intransitive verbs.

- (69) a. [jod] EXIST ‘exist, be located’
 b. [jin] COP, a copula
 c. [ʰdug] ‘be, sit’ (combines functions of copula and existential verbs)

4.11.4.9 Negation

Negative particles immediately precede the words they negate. The particle [ma] is used to negate [jod] and [jin] and past tense forms of other verbs. Before a present tense verb – not an imperative – it forms PROHIBITIONS (PROH):

- (70) a. ma lta
 PROH look(PRS)
 ‘Don’t look!’
- b. lta
 look(IMP)
 ‘Look!’

The particle [mi] negates most adjectives, the existential verb [ʔdug], and present and future forms of other verbs.

The negative particle [ma] coalesces with [jod] and [jin] to form the contracted forms [med] and [min], respectively. The word [med] is also used as a PRIVATIVE suffix – showing a lack of something, in contrast with [tʂan], which shows possession:

- (71) a. dbu-med
 head-NEG.EXIST
 ‘headless’
- b. dbu-tʂan
 head-possessing
 ‘having a head’

[dbumed] and [dbutʂan] are the names of two variants of the Tibetan script.

4.11.4.10 Word order

The basic word order is SOV. As already noted, adjectives usually follow the nouns they modify and adverbs usually precede the verbs they modify.

4.11.4.11 Relative clauses

Relative clauses usually precede their heads and are really nominalized verb phrases that are then linked to the head noun by the genitive case marker. If they follow the head noun, they are not marked in any special way as being relative clauses, but simply end in a nominal particle.

- (72) bgo.ba med-pa=i mi rnam
 clothes NEG.EXIST-NOUN=GEN person PL
 ‘people who have no clothes’

TABLE 4.16

Honorifics in Tibetan

Plain form	Honorific form	Gloss
lus	sku	‘body’
mgo	dbu	‘head’
lag	p ^h ag	‘hand’
mig	sp ^h an	‘eye’
rna	s ^h an	‘ear’
rta	ṭṭ ^h ibs	‘horse’
ⁿdug	bzugs	‘exist’
ṅal.ba	ṅalṅzimpa	‘to sleep’

or, less commonly:

- (73) mi bgo.ba med-pa rnam
 person clothes NEG.EXIST-NOUN PL

4.11.4.12 Honorifics

Although the classical language did not have as elaborate a system of honorifics as Modern Lhasa Tibetan now has, it did have quite an impressive number of honorific words (HON). Most honorifics were nouns, but there were also a few verbs. The honorific word was applied to possessions and actions of persons higher in social status than the speaker, and the plain form was used in informal conversation among friends or to refer to one’s own actions or possessions when speaking to strangers or superiors. Recall a similar situation in Japanese (§4.1.5). More often than not the honorific word was completely different from the plain word (Table 4.16).

4.11.5 SAMPLE TEXT

The sample text below is taken from the third chapter of the biography of the famous Tibetan Buddhist saint Milarepa (མི་ལ་རས་པ [mi.la.ras.pa], 1040–1123 AD). The biography was written by གཙང་སྐྱོན་ཉེ་རུ་ག། [gtsaṅ.sm^hon he.ruka] (1452–1507 AD) and describes in fairly plain language the spiritual journey of Milarepa from being a practitioner of black magic and a murderer to being an enlightened saint. Although the setting is exotic Tibet more than 900 years ago, even those readers who are not at all familiar with Tibet and Tibetan culture will easily understand what Milarepa had to go through on his way to enlightenment: his encounters with such universal human foibles as greed, hatred, indifference, and ignorance in himself and others. In sum, the biography was written with such sincerity and compassion that the reader can easily forget the exotic setting and empathize with Milarepa simply as a fellow human being, who had to contend with many of the same things that we have to go through no matter where or in what century we live.

The setting of the excerpt is as follows. On his deathbed, Milarepa’s father made Milarepa’s paternal aunt and uncle promise to look after his widow and

children. However, after his death, the aunt and uncle broke their promise and robbed the bereaved family of almost all their possessions. Filled with grief, hatred, and yearning for revenge, the betrayed widow hopes that Milarepa will learn black magic and take a terrible revenge on his aunt and uncle. And so Milarepa is sent off to study but, being a young man, soon forgets what it is that he must do. It is this point in the story that the excerpt describes. If you want to learn what happened next, Lhalungpa (1984) gives a good English translation of Milarepa's biography, on which our free translations of this passage are based.

- (74) ང་མི་ཐོད་གདུལ་ན་གློག་སློབ་དུས་རྒྱུ་མཉམ་པོ་ཞིག་བྱུང་བ།
 ŋa mit^hod.gadk^ha na klog slob-dus r̥tsa=i mda=ru
 1SG Mithodgadkha INE reading learning-time Rtsa=GEN lower_valley=ALL
 sk'id-ston=g^{ji} t̥^haŋ-sa t̥^henpo=zig b'uŋ-ba
 happiness-feast=GEN beer-place big=INDF take_place-NOUN

'While I was studying at Mithodgadkha, a big wedding banquet beer celebration took place in the lower valley of Rtsa.'

[slobdus]. One would ordinarily expect [slobpa=i dus] here, that is, a relative clause whose head is the word [dus] 'time': 'at the time I was studying'. Perhaps this is more a compound than a relative clause, and that is why the expected markers are missing.

[b'uŋ-ba]. Verb forms are very often followed by a nominalizing suffix. In many cases it makes little difference whether one considers them to be simply verbs or derived nominals.

- (75) དེའི་གལ་དབུ་ལ་སློབ་དཔོན་སྲུང་དང་ས་པའི་ཕྱག་ཕྱི་ལ་ངས་ཀྱང་ཕྱིན།
 de=i gral-dbu=la slob-dpon sp'an-dranjs-pa=i
 DEM=GEN row-head(HON)=DAT teach-master eye(HON)-invited-NOUN=GEN
 p^hag-p^hi=la ŋa=s k'aŋ pi^hin
 hand(HON)-behind=DAT 1SG=ERG also went

'I went along as the attendant of my teacher, who had been invited to preside over it.'

The sentence seems to split the constituents of the relative clause by placing the head noun in the middle of it. The more usual Tibetan construction would be something like: [dei graldbula sp'andranjspai slobdponla mp^hagsp^hila ŋas kaŋ pi^hin].

[dei graldbula]. Dative case here is used to indicate purpose. [graldbu] is the head of the table, thus, the presider at the feast.

[dranjs] is the past of ['dren].

[p^hagp^hi] 'servant'.

[ŋas]. It is not clear why a subject of what seems to be an intransitive verb ('go') is marked with the ergative case instead of the expected absolutive case.

- (76) དེར་ཆང་མང་བ་དང་སློབ་དཔོན་ལ་ཟུར་ནས་ཀྱང་ཆང་འདྲེན་མང་རབ་བྱུང་བ་ཀྱན་ནས་བཏུངས་པས་བཟི་
 བར་སོང་ཡོད་པ་ལ།
 de=r t̥^haŋ maŋ-ba daŋ slob-dpon=la zur=nas k'aŋ t̥^haŋ ⁿdren
 DEM=ALL beer be_much-NOUN and teach-master=DAT side=ELA also beer serve

marŋ rab b'uŋba kun=nas btuŋs-pa=s bzi-ba=r
 much plentiful happened all=ELA drank-NOUN=ERG inebriated-NOUN=ALL
 soŋ jod-pa=la
 became EXIST-NOUN=DAT

'There was much beer there, and having drunk also of all the beer that was served to him on the side, my teacher became drunk.'

[b'uŋ]. Past of ['b'uŋ].

[btuŋs]. Past of ['ⁿtuŋ]. The sequence verb + nominalizing suffix + ergative marker is another nonfinite verb form. Sometimes it has the force of a subordinate clause of cause ('because'), but most of the time it can be translated into English by a participial clause: 'my teacher having drunk'.

[bzibar]. The allative case here marks the state that is the endpoint of the change.

[soŋ]. Past of ['ⁿgro].

[jodpala]. The dative marker is sometimes used as a conjunction 'and'.

- (77) སློབ་དཔོན་གྱིས་ང་ལ་ཁྱེད་མ་རྒྱུ་རྟེ་ཚོན་ལ་བརྗེས་པས།
 slob-dpon=g'is ŋa=la k'^hos-ma rnam sbskur=te
 teach-master=ERG 1SG=DAT gift-NOUN PL handed_over=SUBORD
 sŋon=la brdzaŋs-pa=s
 ahead=DAT sent_off-NOUN=ERG

'My teacher entrusted his presents to me, sending me ahead.'

[bskur]. Past of [skur].

[brdzaŋs]. Past of [rdzoŋ].

- (78) ཚང་གིས་བཟི་བ་དང།
 tṣ'haŋ=gis bzi-ba daŋ
 beer=ERG drunk-NOUN also

'I was drunk from the beer too.'

[daŋ]. The conjunction 'and' is postpositive, coming after the clause it connects.

- (79) དེའི་ཉིན་སྤོང་ལེན་མ་རྒྱུ་རྟེ་སེམས་འཕྲོས་ནས།
 de=i jin glu-len-mk'^han rnam=la sems 'ⁿros=nas
 DEM=GEN day song-sing-AGT PL=DAT mind continued=ELA

'On that day, my mind dwelt on the singers, so'

- (80) སྤོང་ལེན་སྤོང་འདོད་པ་ཞིག་བྱུང་བ་དང།
 glu-len sŋiŋ-ⁿdod-pa=zig b'uŋ-ba daŋ
 song-sing heart-desire-NOUN=INDF came_about-NOUN and

'I got a desire to sing, too.'

- (81) གྲེ་བ་བདེ་མོ་ཞིག་ཡོད་པས་སྐྱེ་ལེན་ཅིང་འོངས་པས།
 gre.ba bde.mo=zig jod-pa=s glu len=t̪çin ɔŋs-pa=s
 throat good=INDF EXIST-NOUN=ERG song sing=SUBORD came-NOUN=ERG
 ‘Having a good voice, I sang as I came along.’
- (82) རོང་རང་གི་ཁང་པའི་མདུན་ན་ཡར་ལམ་ཡོད་པས་སྒྲིའི་ཐང་དུ་སྐྱབས་རུང་སྐྱེ་ལེངས་པས།
 ŋed raŋ=gi kʰaŋ-pa=i mdun=na jar-lam jod-pa=s
 1PL.PRO self=GEN house-NOUN=GEN front=INE upper-road EXIST-NOUN=ERG
 sgo=i tʰad=du slebs ruŋ glu blaŋs-pa=s
 door=GEN forward=ALL arrived though song sang-NOUN=ERG
 ‘The road ran in front of our house, and when I arrived at the door,
 I was still singing a song.’
 [blaŋs]. Past of [len].
- (83) ཇམ་མ་ནང་ན་ཡོས་རྫོད་ཀྱི་ན་ཡོད་པས་ཐོས་ནས།
 ʔa.ma naŋ=na jos ŋɔd=kʰin jod-pa=s tʰos=na
 mother inside=INE roasted_barley roast=PROG EXIST-NOUN=ERG hear=ELA
 ‘My mother was inside roasting barley, and she heard me.’
 [ŋɔd kʰin jod]. Placing [kʰin jod] after the verb seems to indicate a continuing action that
 is taking place at this very moment in the present or at some particular moment in the past,
 simultaneous with some other event: ‘Mother was just then in the process of parching barley.’
 [tʰos]. Some verbs have inflected forms identical with that of the present tense form. Thus,
 [tʰos] can be either present or past tense, depending on the context.
- (84) ཅི་ཟེར་སྐད་འདི་ནི་ངའི་བུའི་སྐད་འདྲ། རོང་མ་སྐད་ཀྱི་སྐྱུག་པ་ནི་ས་ཐོག་ན་མེད་པས་
 t̪çi zer skad ʰdi ni ŋa=i bu=i skad ʰdra ŋed ma-smad=kʰi
 what(Q) say voice this TOP 1SG=GEN SON=GEN voice similar 1PL mother-child=GEN
 sdug-pa ni sa tʰog=na med-pa=s
 misery-NOUN TOP earth top=INE NEG.EXIST-NOUN=ERG
 ‘“What?” she said. “This voice is like my son’s voice. Our misery, mother and
 children, there’s nothing on earth . . .”’
 This appears to be a kind of truncated comparative construction: ‘there is no suffering on
 earth (greater than that of) us, mother and children’. Milarepa’s mother keeps using this
 expression as a kind of formula; thus it is no wonder that the comparison marking has been
 elided for brevity.
 [zer]. This verb is used to signal a quotation. Sometimes it comes at the end of the quote,
 and sometimes it is embedded within a quotation, often as a second element. It is sometimes
 very difficult to figure out exactly where a quotation begins and ends, since there are no
 punctuation marks to signal its boundaries.
 [ni]. This topic marker adds a certain degree of emphasis to a noun phrase or a clause after
 which it appears; it sometimes seems to have a contrastive function as well.

- (85) ཁོ་ནི་སྤྱི་ལེན་མི་ཐད་སྐྱུ་ནས་ཡིད་མ་ཆེས་པར་བློས་པས།

k^ho ni glu len mi t^had s^hnam=na^s jid ma t^hes-pa=r
 3SG TOP song sing NEG right think=ELA mind NEG believe-NOUN=ALL
 bltas-pa=s
 looked-NOUN=ERG

'Thinking it would not be right for him to be singing a song, in disbelief she looked.'
 [bltas.pas]. The allative case here forms a kind of manner adverbial phrase: 'unbelievingly'.

- (86) ང་ཡིན་པར་ཤེས་ནས་ཉ་ལས་ཏེ། སྐྱམ་པ་གཡམས་སྤྱོད་པ་ཡོས་དུག་གསལ་གཡོན་དུ་བོར།

ŋa jin-pa=r ces=na^s ha.las=te skam-pa gjas=su
 1SG be-NOUN=ALL perceive=ELA be_shocked=SUBORD tongs-NOUN right=ALL
 bor jos-dkrugs gjon=du bor
 threw barley-stirrer left=ALL threw

'Perceiving that it was me, she was shocked, and she threw the tongs to the right and she threw the barley-stirrer to the left.'

[ŋa jinpar]. The allative case marker is also used to mark a noun phrase or subordinate relative clause that is the direct object of a verb of knowing, perceiving, etc.

[gjasu; gjondu]. Here the allative case indicates the direction of a motion.

[bor]. Past of ['ⁿbor].

- (87) ཡོས་རྫོང་འཕྲོ་འཆོག་རྩ་བརྩལ་ལག་པ་གཡམས་སྤྱོད་པ་ཞིག་བྱིར། གཡོན་དུ་ཐལ་བ་སྐར་གང་བྱིར།

jos ŋod ⁿp^hro ⁿts^hig=tu bt^hug lag-pa gjas=su
 barley parch continue burn=ALL let(PST) hand-NOUN right=ALL
 jog-pa=zig k^her gjon=du t^hal-ba spar-gaŋ k^her
 poker-NOUN=INDF took left=ALL ashes-NOUN handful-measure took

'She allowed the parched barley to continue to burn and took a poker in her right hand and a handful of ashes in her left.'

Parched barley, as well as other parched grains, is the staple food of Tibetans.

[ⁿts^hig tu bt^hug]. The allative case marker is used here to connect the two verbs.

[bt^hug]. Past of ['ⁿbt^hug].

[k^her]. Past of ['ⁿk^her].

- (88) སྐས་རིང་བ་ལ་བབས། ཐུང་བ་ལ་མཚོངས་ནས་ཕྱིར་ཐོན་བྱུང་གླེ།

skas riŋba=la babs t^huŋba=la mt^hoŋs=na^s p^hi=r
 step steep=DAT descended low=DAT leaped_over=ELA outside=ALL
 t^hon b^huŋ=ste
 appear became=SUBORD

[babs]. Past of ['ⁿbab].

[b^huŋ]. Past of ['ⁿb^huŋ]. This verb is used as a kind of past tense auxiliary verb, especially in those cases when the main verb has no special past tense form different from the present tense form.

- (89) ཐལ་བ་དེ་གདོང་ལ་བཏབ། ཡོག་པ་དེ་མགོ་ལ་ལན་འགའ་བརྒྱབས་ནས།
 t^hal-ba de gdoŋ=la btab jogpa de mgo=la lan ^oga brg^jabs=nas
 ashes-NOUN DEM face=DAT threw poker DEM head=DAT time few hit(PST)=ELA
 ‘She threw those ashes in my face and hit me a few times on the head with that poker.’
 [btab]. Past tense of [gtoŋ].

[brg^jabs]. Past tense of [rg^jab].

[mgo la ... brg^jabs]. With some verbs [la] = DAT is used to mark a direct object. This is somewhat parallel to such English verb phrases as *look at something* in which the object of the verb *look* is marked by the preposition *at*.

- (90) ཡབ་མི་ལ་ཤེས་རབ་རྒྱལ་མཚན་བྱིད་ལ་བུ་འདི་འདྲ་སྐྱེས་སོ།
 jab mi.la ཅེས་ལ་རྒྱལ་མཚན་གྱི་ལས་སྐྱོས་ལ་གཟིགས་དང་ཟེར་ཨ་མ་ཡུག་ཅིག་བརྒྱལ་ནས་འབྱེལ་
 father(HON) Mila Wisdom victory-sign you=DAT son this-like was_born=FP
 ‘‘Father Mila Banner-of-Wisdom, this is the kind of son that was born to you!’’

- (91) བྱིད་རབས་ཆད་གདའ་བ། དེད་མ་སྤང་གྱི་ལས་སྐྱོས་ལ་གཟིགས་དང་ཟེར་ཨ་མ་ཡུག་ཅིག་བརྒྱལ་ནས་འབྱེལ་
 འདུག་པའི་ཚེ།

k^hed rabs-འདྲ་ad gda-ba ŋed ma-smad=kⁱ lasskos=la gzigs danj
 2 lineage-broken be-NOUN 1PL mother-children=GEN fate=DAT look(HON) IMP
 zer །a.ma jug=འདྲིག brg^jal=nas ^og^jel ^odug-pa=i འཕྲིལ་
 say mother moment=INDF faint=ELA overcome EXIST-NOUN=GEN time

‘‘You have a broken lineage! Look at our fate, mother and children!’’ said mother, whereupon she fainted.’

[danj]. After a verb, [danj], which usually means ‘and’, is sometimes used to signal the imperative mood, especially in the case of verbs that do not have an imperative form different from the present tense form.

[འཕྲིལ་]. Here we have an example of a relative clause, whose head noun is ‘time’, that acts as an adverbial subordinate clause of time: ‘when’, ‘at the time of’.

- (92) སྲིད་མོ་ཡང་ཐོན་བྱུང་སྟེ། ཨ་ཚོའི་བསམ་སྒྲོལ་ཨ་མ་ལ་ལྟོས་དང་ཟེར་བུ་ཞིང་འདུག་པ་ལ། བདེན་སྒྲུམ་ནས།
 srijmo jan t^hon b^huŋ=ste adz̄o=i bsam-blo=la
 sister(male’s) also came_out became=SUBORD elder_brother=GEN thought-mind=DAT

།a.ma=la ltos danj zer ŋu=མྱེན ^odug-pa=la bden-s^ham=nas
 mother=DAT look(IMP) IMP say weep=SUBORD be-NOUN=DAT true-think=ELA

‘My sister also came out. ‘‘Big brother, consider your thinking, look at mother!’’, she said, weeping, and because of this right thinking.’

[t^hon]. Past of [t^hon].

[b^huŋ]. Past of [b^huŋ].

[ltos]. Imperative of [lta].

[ŋu=མྱེན ^odug-pa]. The verb [dug] often appears with [མྱེན] ([འདྲིག]) as a sort of progressive aspect auxiliary.

- (93) ང་རང་ཡང་མཆི་མ་མང་པོ་ཤོར།
 ŋa-raŋ jaŋ m̥t̥ʰi-ma maŋpo ɕor
 1SG-self also tear-NOUN much fell
 ‘I too shed many tears.’
 [ɕor]. Past of [ʰt̥ɕor].
- (94) དེད་མིང་སྲིད་གཉིས་ཀྱིས་བུ་བའི་ངང་ནས་ཨ་མའི་ལག་པ་མཉེད་ཅིང་བོས་པས་དར་ཅིག་ནས་བརྒྱལ་བ་
 སངས་ཉེ་ལངས་བྱུང་ནས།
 ŋed miŋ-sriŋ gnis=kʰis ŋu-ba=i ŋaŋ=naŋ ʔa.ma=i lag-pa
 1PL brother-sister two=ERG weep-NOUN=GEN state=ELA mother=GEN hand-NOUN
 m̥ŋed=t̥ɕiŋ bos.pas dart̥ɕiŋ nas br̥gʷal-ba
 rub_between_hands=SUBORD called_name a_little_while ELA faint-NOUN
 saŋs=te laŋs bʰuŋ=naŋ
 dispelled=SUBORD stood_up became=ELA
 ‘We two, brother and sister, weeping, rubbed mother’s hands, calling her name.
 After a little while she came to and stood up.’
 [bos]. Past of [ʰbod].
 [laŋs]. Past of [laŋ].
 [bʰuŋ]. Past of [mʰuŋ].
- (95) བུ་བའི་གདོང་གིས་ང་ལ་ཅེ་རེ་བལྟས་ནས།
 ŋu-ba=i gdoŋ=gis ŋa=la t̥ɕe.re bltas=naŋ
 weep-NOUN=GEN face=ERG 1SG=DAT stare looked=ELA
 ‘She looked at me fixedly with her weeping face.’
 [bltas]. Past of [lta].
- (96) བུ་རང་རེ་མ་སྐད་པས་སྐྱབ་པ་ས་ཐོག་ན་མེད་པ་ལ་བྱོད་སྐྱེ་ལེན་པ་དྲན་པ་རང་འདུག་གས།
 bu raŋre ma-smad-pa=s sdug-pa sa-tʰog=na
 son 1PL mother-children-NOUN=ERG misery-NOUN earth-top=INE
 med-pa=la kʰod glu len-pa dren-pa raŋ ʱdug=gam
 NEG.EXIST-NOUN=DAT 2 song sing-NOUN think_of-NOUN even EXIST=Q
 ‘“Son, as the likes of our misery, mother and children, are not found anywhere in the
 world, how do you even think of singing songs?”’
- (97) ང་མ་རྒན་མོ་ནི་བསམ་མཚོན་གཏོང་གིན་ཡི་ཕྱག་པ་དང་བྱོལ་བ་མིན་པ་མི་འདུག་ཟེས།
 ŋa ma r̥ganmo ni bsammno gtoŋ=gin ji.mug-pa daŋ ŋu bro.ba
 1SG mother old TOP thinking let=SUBORD despair-NOUN and weep wish
 min-pa mi ʱdug zer
 NEG.COP-NOUN NEG EXIST say

“As for me, your old mother, when I think of it, I have nothing but despair and a need to cry”, she said.’

[gin] is a form of [k’in], marking the preceding as a subordinated present tense clause.

The force of the two negated verbs in a row (‘it is not + it exists not’) seems to be ‘if not X, there is nothing else but’: ‘I can only.’

- (98) མྱེ་ལྷགས་འདོན་ཞིང་དེད་མ་སྣེད་གསུམ་ཀས་ངུས་སོ།།
 smre.sɲags ʰdon=ziŋ ŋed ma-smad gsum-ka=s ŋus=so
 lamentation utter=SUBORD 1PL mother-children three-NOUN=ERG weep(PST)=FP
 ‘Uttering lamentations, the three of us, mother and children, wept.’
 [ŋus]. Past of [ŋu].

4.12 Exercises

4.12.1 COMPARISON OF MODERN LHASA TIBETAN AND OLD TIBETAN

Compare the Old Tibetan (OTB) Tibetan words in Table 4.17 and their pronunciation in one version of Modern Lhasa Tibetan (Zhōngyāng Mǐnzú Xuéyuàn 1987). The Old Tibetan is essentially a transcription of the classical orthography. What general sound changes have occurred between that stage and the stage of Modern Lhasa Tibetan? How can the tones of the latter be predicted from the spelling? In order to answer the latter question you will have to take into consideration a number of factors, such as the quality of syllable-initial consonants, vowel length, and even the silent letters, that is, letters that are no longer pronounced.

4.12.2 COMPARISON OF TIBETAN AND CHINESE

Classical Tibetan and Mandarin Chinese are considered to be members of the same language family, Sino-Tibetan. However, that does not necessarily mean that they are very similar in most respects. After reading the sketches of these two languages (§4.11 and §4.10), compile a list of major differences and similarities between the two. For example, Classical Tibetan, as originally pronounced, was not a tone language, whereas Mandarin Chinese is. In both, however, plural number is only optionally marked on nouns, whereas it is obligatory in the pronouns. Do the differences outnumber the similarities or vice versa? Don’t include language universal features in your list of similarities. For example, it is not relevant to this exercise that both languages have vowels.

4.12.3 GRAMMATICAL RELATION MARKING IN VARIOUS LANGUAGES

Compare and contrast how different grammatical relations are signaled in Mandarin Chinese, Classical Tibetan, Russian, and English. You may refer either

TABLE 4.17

Old Tibetan and Modern Lhasa Tibetan

	Old Tibetan	Modern Tibetan	Gloss
1.	k ^h a	k ^h aḷ	‘mouth’
2.	sa	saḷ	‘earth’
3.	rigs	ṛiḷ	‘race, lineage’
4.	sems	semḷ	‘heart, mind’
5.	ḡa	ḡaḷ	‘flesh’
6.	ḡar	ḡa:ḷ	‘east’
7.	ṭḡ ^h aŋ	ṭḡ ^h aŋḷ	‘beer’
8.	k ^h oŋ	k ^h oŋḷ	‘he, she’
9.	gsum	sumḷ	‘three’
10.	stoŋ	toŋḷ	‘thousand’
11.	nub	nuḷ	‘west’
12.	bod	pøḷ	‘Tibet’
13.	p ^h ag	p ^h aḷ	‘pig’
14.	mdun	tṽ:ḷ	‘front’
15.	riŋ	ṛiŋḷ	‘long’
16.	za	saḷ	‘eat’
17.	mduŋ	tuŋḷ	‘spear’
18.	sgam	kamḷ	‘box, chest’
19.	lags	laḷ	‘yes’
20.	zas-pa	seḷ-pa	‘food’
21.	ḷṭḡags	ṭḡaḷ	‘iron’
22.	sbrul	ty:ḷ	‘snake’
23.	dkar-po	ka:l-po	‘white’
24.	mḍzes	ṭseḷ	‘beautiful’
25.	ḷa	ḷaḷ	‘god’
26.	la	laḷ	‘mountain pass’

to the sketches in this chapter and the previous chapter or refer to the more detailed descriptions cited in the Suggested Readings lists.

4.12.4 HMONG LANGUAGES AND CHINESE

Examine the data in Table 4.18 from three different Hmong languages and Mandarin Chinese (Wáng 1985) and then answer the following questions:

- On the basis of this data, can it be shown that the three Hmong languages are genetically related to each other?
- On the basis of this data, can it be shown that Mandarin Chinese and the Hmong languages are related?
- Does there seem to be borrowing between Hmong and Mandarin? If so, in which direction did the borrowing likely go? Please note that we have cited Mandarin words in their standard form, but that Hmong speakers

TABLE 4.18

Hmong and Mandarin Chinese

	Hmong A	Hmong B	Hmong C	Mandarin	Gloss
1.	u ¹	ə ¹	tɬe ¹	ʃwəi ¹	'water'
2.	pu ¹	pi ¹	pe ¹	san ¹	'three'
3.	p ¹ jei ¹	ɭu ¹	plou ¹	si ¹	'four'
4.	ji ¹	ɯa ¹	ji ¹	pa ¹	'eight'
5.	pa ¹	pa ¹	pua ¹	pai ¹	'hundred'
6.	ⁿ pe ¹	pe ¹	ⁿ po ¹	ɕəpə ¹	'snow'
7.	noŋ ¹	noŋ ¹	naŋ ¹	y ¹	'rain'
8.	pə ¹	pi ¹	pu ¹	ʃwəi ¹	'sleep'
9.	ta ⁿ pa ¹	pa ¹	ⁿ pua ¹	ʃu ¹	'pig'
10.	jin ¹ haŋ ¹	ɯen ¹ haŋ ¹	ɯen ¹ haŋ ¹	in ¹ xaŋ ¹	'(financial) bank'
11.	ʔɕi ¹ pa ¹	ti ¹ pa ¹	ti ¹ pa ¹	ti ¹ pa ¹	'eighth'
12.	ŋhe ¹	ŋhe ¹	ŋo ¹	ɿi ¹	'sun'
13.	noŋ ¹	naŋ ¹	nau ¹	ʔ ⁿ i ¹	'eat'
14.	hu ¹	fu ¹	fu ¹	xu ¹	'lake'
15.	pei ¹	paŋ ¹	paŋ ¹	xwa ¹	'flower'
16.	me ¹ ʃei ¹	me ¹ ʃei ¹	me ¹ ʃuei ¹	mwə ¹ ʃwəi ¹	'ink'
17.	moŋ ¹	moŋ ¹	mo ¹	k ^h y ¹	'go'
18.	to ¹	to ¹	to ¹	ʃən ¹	'deep'
19.	le ¹	le ¹	lo ¹	twan ¹	'short'
20.	ⁿ tu ¹	ta ¹	ⁿ te ¹	ʔ ⁿ aŋ ¹	'long'
21.	joŋ ¹ ji ¹	ɣu ¹ le ¹	joŋ ¹ ji ¹	ɿuŋ ¹ i ¹	'easy'
22.	ji ¹	ji ¹	ji ¹	i ¹	'100,000,000'
23.	a ¹	i ¹	i ¹	i ¹	'one'
24.	ʔɕi ¹ ji ¹	ti ¹ ji ¹	ti ¹ ji ¹	ti ¹ vi ¹	'first'
25.	ʔɕi ¹ ʔwu ¹	ti ¹ ʔvu ¹	ti ¹ ʔwu ¹	ti ¹ ʔu ¹	'fifth'

have also been in contact with speakers of Southwestern Mandarin, which is similar to but not identical with the standard.

In item 16, the Mandarin Chinese word for 'ink' is morphologically complex: the first syllable means 'India ink stick', and the second means 'water'. Chinese writers used ink sticks that they partially dissolved by rubbing them on an ink stone on which they poured a bit of water whenever they needed ink for writing with the traditional Chinese writing brush.

4.12.5 HONORIFICS, ELLIPSIS, AND PRONOUNS IN JAPANESE

Japanese allows much more ELLIPSIS than English, omitting words that one might logically expect. It does not require DUMMY PRONOUNS – pronouns that don't refer to anything but are required by the syntax, such as in English *it* in sentences like *It is raining*. Personal pronouns are used very sparingly. Japanese possesses a highly

- (4) Mrs. Nakada:

háí | açítá wa çúðzín no háha no tandzoo-bí
 yes tomorrow TOP husband(NHON) GEN mother(NHON) GEN birth-day
 des-u node ||
 COP.POLITE-IPFV because

NHON = nonhonorific: the word is not honorific, but there is an honorific counterpart that it contrasts with.

- (5) Mrs. Tanaka:

ðza | nápnika pulézento o o-ka-i ni nañi-más-u no ||
 well Some present ACC HON-buy-INF DAT become-POLITE-IPFV Q
 [ni nañimásu] after an honorific infinitive is a way of honoring the person who is the agent of the verb.

[no] is an interrogative particle used especially by women.

- (6) Mrs. Nakada:

háí | depáato e maiñi-más-u ||
 yes department_store to go(HBL)-POLITE-IPFV

- (7) nápnika í-i áidea ga át-tala oçie-te kudasá-i ||
-
- some good-IPFV idea NOM exist-COND teach-GER give(to_inferior)-IMP

[íi]. Many Japanese adjectives inflect in ways analogous to verbs, here taking an imperfective aspect.

[áttala]. Conditional mood, 'if'.

[kudasái] is used with a gerund to make a polite request; it can be translated as 'please'.

- (8) Mrs. Tanaka:

dó-nna monó ga o-sucí des-u ka ||
 Q-kind_of thing NOM HON-liked COP.POLITE-IPFV Q

[sucí]. Many other adjectives behave much like nouns. This adjective-noun is syntactically the argument of the copula, 'a thing is liked', but the English idiom would turn this around to form a verb: 'likes a thing'.

- (9) Mrs. Nakada:

ci-lu monó jóñi mo tabé-lu monó ga sucí na n des-u ||
 wear-IPFV thing than even eat-IPFV thing NOM liked COP PRO COP.POLITE-IPFV
 [ci-lu monó]. Clauses can be placed before nouns to modify them.

This illustrates the Japanese comparative construction: X jori (mo) Y + subject marker + sucí na = 'likes Y more than X'.

- (10) Mrs. Tanaka:

ðzáa | kéeci o saçiage-tála ikága deç-óo ka ||
 well_then cake ACC give(to_superior)-COND how COP.POLITE-TENTATIVE Q

- (11) Mrs. Nakada:

démo ɕúdz̄in ga kéeci o kat-te ci-más-u ||
 but husband NOM cake ACC buy-GER come-POLITE-IPFV

- (12) Mrs. Tanaka:

o-káa-sama wa t̄ɕokoléeto ga o-sucí des-u ka ||
 HON-mother-HON TOP chocolate NOM HON-liked COP.POLITE-IPFV Q

- (13) Mrs. Nakada:

há | dái-sucí des-u kedo t̄ɕokoléeto wa musumé ga
 yes much-like COP.POLITE-IPFV although chocolate TOP daughter(NHON)NOM
 kai-más-u ||
 buy-POLITE-IPFV

- (14) honto ni || náni o age-tála ii n deç-óo
-
- real DAT what ACC give(to_superior)-COND good-IPFV PRO COP.POLITE-TENTATIVE
-
- ne ||
-
- Q(TAG)
-
- [honto ni] 'really'.

- (15) o-taku no go-ɕúdz̄in no o-káasama wa dó-nna monó ga o-sucí
-
- HON-family GEN HON-husband GEN HON-mother TOP Q-kind_of thing NOM HON-liked
-
- des-u ka ||
-
- COP.POLITE-IPFV Q

- (16) Mrs. Tanaka:

ut̄ɕi wa nándemo sucí des-u kala koojuu mondai wa náí
 family(NHON)TOP anything liked COP.POLITE-IPFV because such problem TOP not
 des-u ne ||
 COP.POLITE-IPFV Q(TAG)

- (17) go-ɕúdz̄in to od̄z̄óosan wa mai-toçi onad̄zi monó o pulézentó
-
- HON-husband and daughter(HON) TOP every-year same thing ACC present
-
- nasál-u n des-u ka ||
-
- do(HON)-IPFV PRO COP.POLITE-IPFV Q

Most loanwords in Japanese are nouns; however, most of these can be easily made into verbs by the addition of [suru] 'to do'. Thus [pulézentó] 'a present' can be made into a verb 'to make a gift or a present': [purézentó suru]. Here [pulézentó nasálu] is an honorific equivalent, because [nasálu] is the honorific equivalent of the plain verb [sulu]. The humilific equivalent of [sulu] is [itásu].

(18) Mrs. Nakada:

háí | só-o des-u ||
yes MEDIAL-ADV COP.POLITE-IPFV

Medial demonstratives refer to things close to the hearer.

(19) déwa | tónikaku it-te m̃i-maç-óo ||

well anyway go-GER try-POLITE-TENTATIVE

4.13 Suggested readings

4.13.1 GENERAL

- ❑ Far Eastern languages (Egerod 1991). Credible account of the genetic connections among the region's languages and of their typological features.
- ❑ *The languages of East and Southeast Asia: An introduction* (Goddard 2005). Basic introduction to the languages and language groups of East and Southeast Asia. It covers phonology, syntax, typology, and writing systems. In the back are exercises for each chapter.
- ❑ *The languages of Japan and Korea* (Tranter 2012). Chapters by experts covering the historical development of Korean and Japanese, along with descriptions of the modern dialects of Korean, Japanese, and Ryukyuan.

4.13.2 ALTAIC AREA

- ❑ *The genetic relationship of the Ainu language* (Patrie 1982). Concludes that Ainu is an Altaic language, contrary to most scholars.
- ❑ *Japanese and what other Altaic languages?* (Unger 1990). A critical attack on the Altaic hypothesis, especially on the inclusion of Japanese in Altaic.
- ❑ *The Korean language* (Sohn 2001). Introduction to the linguistic structure and historical development of Korean.
- ❑ *The languages of Japan* (Shibatani 1990). Good introduction to Japanese and Ainu.
- ❑ *The Mongolic languages* (Janhunen 2003). Scholarly contributions by experts add up to a thorough history of Mongolic, with sketches of important languages of the group and a chapter comparing Mongolic with Turkic.
- ❑ *The Turkic languages* (Johanson & Csató 2006). An exhaustive collection of chapters by experts on Turkic languages as a whole, plus sketches of over a dozen individual Turkic languages.

4.13.3 PALEOSIBERIAN AREA

- ✘ Paleosiberian and other languages (Comrie 1981). Presents interesting linguistic features of the half dozen small groups and isolates comprising this collection of languages that have resisted classification.

4.13.4 SINO-TIBETAN

- ✘ *Chinese* (Norman 1988). The history and linguistic structure of Chinese aimed at a slightly more advanced level than Sun (2006).
- ✘ *Chinese: A linguistic introduction* (Sun 2006). An entry-level account of Chinese in its historical and cultural context.
- ✘ *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction* (Matisoff 2003). Complete reconstruction of the sounds and suffixes of Proto-Tibeto-Burman.
- ✘ *The Sino-Tibetan languages* (Thurgood & LaPolla 2003). A collection of articles by specialists, with several chapters on the family as a whole, several more on Chinese, and 32 chapters on individual Tibeto-Burman languages.
- ✘ Sino-Tibetan linguistics: Present state and future prospects (Matisoff 1991). An excellent overview of the Sino-Tibetan family of languages including languages only tentatively assigned in the family, with typological information about all of the groups.

4.13.5 HMONG-MIEN

- ✘ *Hmong-Mien language history* (Ratliff 2010). A reconstruction of Proto-Hmong-Mien phonology and morphology, including tone.

4.13.6 TAI-KADAI

- ✘ *The Tai-Kadai languages* (Diller et al. 2011). Twenty substantial sketches of languages of this group by specialists plus an examination of hypotheses for connections of Tai-Kadai to possible larger groups Austro-Tai, Sino-Tai, and Sino-Tibetan-Austronesian.

4.13.7 AUSTROASIATIC

- ✘ *The Munda languages* (Anderson 2008). Fourteen chapters by experts covering phonology, morphology, syntax, semantics, and lexicon in languages of this group.
- ✘ *Vietnamese* (Nguyễn 1997). Vietnamese is not a typical Austroasiatic language due to heavy influence from Chinese.

4.13.8 DRAVIDIAN

- ✎ *The Dravidian languages* (Krishnamurti 2003). Detailed survey of the family by an eminent specialist, including history, structure, writing systems, and unwritten languages.

Africa

This chapter locates Africa's languages within their ancestral families and highlights a variety of linguistic features that set these languages apart from one another and from other languages around the world.

Of the world's 7,358 languages, over 2,000 are indigenous to Africa. Surprisingly, the analysis of Greenberg (1963) divided this large number of African languages into just four families: Afro-Asiatic, Nilo-Saharan, Niger-Congo, and Khoisan (Table 5.1). Nowadays, most linguists believe that this consolidation was a little too ambitious, particularly as applied to the very diverse Khoisan languages, but Greenberg's analysis is still the most recognized framework for discussing Africa. As usual, we discuss in other chapters language families that are widely spoken in Africa but that lie mostly outside of the continent: in particular, Austronesian and Indo-European (§5.5).

In terms of number of speakers, the Niger-Congo and the Afro-Asiatic families are each larger than all but two of the world's language families. Those two non-African families are Indo-European (§3.1) and Sino-Tibetan (§4.3), which together are spoken by two thirds of the world's population.

Scholars came late to the study of African languages. An early landmark was Koelle (1854), a set of nearly 300 concepts translated into over 100 different African languages. The book was compiled mainly with the help of freed slaves in Freetown, the capital of present-day Sierra Leone. Koelle's word lists are accompanied by

TABLE 5.1
Languages of Africa

Name	Section	Size	Location
Afro-Asiatic	§5.1	9	north; Middle East
Nilo-Saharan	§5.2	8	central
Niger-Congo	§5.3	9	south
Khoisan area	§5.4	6	southwest
Other languages	§5.5	8	

groupings of the languages based on shared similarities. Some of Koelle's groupings correspond remarkably well to the accepted groupings we note in this chapter.

The study of African languages has opened the door to many new linguistic discoveries. Research on these languages in recent decades has revised and overturned many earlier notions about fundamental linguistic concepts. The list includes linguistic tone; interactions between phonology, morphology, and syntax; the inventory of sounds in language; and a variety of syntactic phenomena. One unfamiliar but worthwhile example of a syntactic phenomenon is LOGOPHORICITY, a type of pronominal reference requiring an extra set of pronouns in some African languages (§5.3.5). Another is the serial verb (§5.3.5), which, as we have noted, is also common in Asia (§4.10.3.8).

In Africa, as elsewhere, one of the greatest linguistic concerns is how many languages of this continent will continue to be spoken by future generations. *Ethnologue* lists 46 African languages as nearly extinct. Many others are in danger of disappearing in a generation or two.

5.1 Afro-Asiatic

The name of the Afro-Asiatic family reflects its location, which includes both Africa and the Middle East. In Africa, these languages are found in North Africa, East Africa, and in a relatively narrow belt just south of the Sahara desert (see the languages labeled with a raised “1” in Figure 5.1). Among the most widely known Afro-Asiatic languages are Arabic, Hebrew, Somali, Amharic, Berber, and Hausa. The roughly 400 languages of this family fall into six groups, described in the subsections listed in Table 5.2, though there are some controversies as to which languages belong to which groups.

5.1.1 SEMITIC

The languages belonging to the Semitic branch of Afro-Asiatic cover most of North Africa, all of the Arabian peninsula, Iraq, Israel, Lebanon, Syria, and much of Ethiopia and Eritrea, as well as the island of Malta and parts of Iran and Turkey. The branch is named after Noah's son Shem. The Bible named him as the progenitor of Jews and of people of several other nations, most of whom speak languages related to Hebrew. The word *Semitic* begins with [s] instead of [ʃ], because it was Latinized in the nineteenth century, and Latin has no [ʃ].

Table 5.3 lists a sampling of the Semitic languages, following the subgrouping of Rubin (2008).

The earliest attested Semitic language was Akkadian, which was first written around 2600 BC using Sumerian cuneiform. It and its second-millennium dialects, Assyrian and Babylonian, were important LINGUA FRANCA in the ancient Near

TABLE 5.2

Some Afro-Asiatic languages

Name	Section	Size	Location
Semitic	§5.1.1	9	North, Horn, W Asia
Berber	§5.1.2	8	North
Chadic	§5.1.3	8	Central Sahel
Cushitic	§5.1.4	8	Nile, Horn
Omotic	§5.1.5	7	Horn
Egyptian	§5.1.6	0	Nile

TABLE 5.3

Some Semitic languages

Name	Size	Location
East Semitic		
· Akkadian (AKK)	0	Iraq
West Semitic		
· Central Semitic		
· · Northwest Semitic		
· · · Canaanite		
· · · · Hebrew (HEB)	7	Israel
· · · · Phoenician (PHN)	0	Lebanon
· · · Aramaic (ARC)	6	Iraq
· · · · Syriac (SYC)	0	Iraq
· · · · Turoyo (TRU)	5	Turkey
· · · · Western Aramaic (AMW)	5	Syria
· · · · Northeastern	6	Iraq
· · · · Mandaic (MID)	4	Iran
· · Arabic (ARA)	9	Egypt
· · · Maltese (MLT)	6	Malta
· Ethiopic		
· · North Ethiopic		
· · · Ge'ez (GEZ)	0	Eritrea
· · · Tigrinya (TIR)	7	Eritrea
· · · Tigré (TIG)	7	Eritrea
· · South Ethiopic		
· · · Amharic (AMH)	8	Ethiopia
· · · East Gurage (STV)	5	Ethiopia
· · · Central West Gurage (SGW)	6	Ethiopia
· Modern South Arabian		
· · Mehri (GDQ)	6	Yemen

East. A lingua franca is a language widely learned as a second language, enabling communication across a wide area where many languages are spoken. Akkadian probably ceased to be anybody's native language by around 300 BC.

The main Central Semitic languages are Aramaic, Arabic, and Hebrew, along with a few ancient languages such as Phoenician. Most of the Central Semitic languages have a long written history.

Aramaic was the successor to Akkadian as a major lingua franca in the Near East, especially under the Persian Empire. Its importance during the development of Judaism and Christianity has made it an important language of religion. Notable examples are the Aramaic of the Gemara, which are the later portions of the Talmud; and Syriac, which was a main vehicle for the spread of Christianity in the Near East and remains the liturgical language of several eastern Christian churches. Today it survives in the form of 19 different languages or dialects spoken in scattered communities in Syria, Turkey, Iraq, Iran, and Israel.

Arabic is by far the most widely spoken of all the Afro-Asiatic languages. It will be discussed more fully in the sketch of Modern Standard Arabic (§5.6).

The Canaanite languages were attested in the Levant in the second and first millennia BC. The best known of these are Hebrew and its close relative, Phoenician. Phoenician has already been mentioned as the language of the people who were instrumental in cultivating the old Canaanite script and passing it on to the Greeks and many other cultures (§2.2.2). Hebrew is well known as the language of Jews and Samaritans. All of the Canaanite languages became extinct as a language of daily communication by about the first century AD, but Hebrew has always been maintained as a religious language of Judaism and was revived as an everyday language in the twentieth century with the creation of the state of Israel.

The last group of languages to consider here are the South Semitic languages, which are spoken primarily in Ethiopia and Eritrea and across the Red Sea in the Arabian peninsula. Ge'ez is the classical South Semitic language and was used as a literary language long after it ceased to be spoken. Nowadays, it is still used as a liturgical language by the Ethiopian Orthodox Church as well as by the Beta Israel Jewish community, most of whom now live in Israel. Ge'ez has its own derivative of the Canaanite script. Originally it was an abjad, but it developed diacritic modifications to represent vowels, making it an abugida. The two most widely known modern languages from this group are Tigrinya, which is especially widespread in Eritrea, and Amharic, which is especially strong in Ethiopia; both use the Ge'ez script.

Semitic languages spoken in the southern Arabian peninsula are called South Arabian languages. Though these languages are related to Arabic, as are all Semitic languages, they are not dialects of Arabic, and so must not be called **South Arabic*. Mehri is the most widely spoken of them, with speakers mostly in Yemen and Oman.

Semitic languages are particularly valuable to historical linguists because of the fairly great time depth at which they are attested. Proto-Semitic is notable for having ejective consonants, which we met earlier in the Caucasus (§3.3.3). It is reconstructed as having 14 contrasting fricatives:

(1) Proto-Semitic fricatives

dental	alveolar	postalveolar	uvular	pharyngeal	glottal
θ	s	ʃ	χ	ħ	h
θ'	s'				
ð	z		ʁ	ʕ	
	ʔ				
	ʔ'				

Not only is this a very large number of fricatives, but some of them are fairly infrequent in the languages of the world. In particular, pharyngeal fricatives are quite rare, and it may surprise English speakers to learn that the dental fricatives [θ] and [ð] are almost as rare. Ejective fricatives are very uncommon, too, much more so than ejective stops. The sound [ʔ], by the way, is a voiceless lateral fricative. It can be made by putting the tongue in the position to make an [l], but instead making a hissing sound by forcing air between one side of your tongue and your teeth.

Several of the Semitic languages, such as Arabic, have replaced the ejectives with PHARYNGEALIZED CONSONANTS. Pharyngealized consonants like [tˤ] are made with the same airflow as [t], with air coming from the lungs, but with the throat constricted, as for the pharyngeal consonants. The term EMPHATIC CONSONANT is a cover term that can refer either to ejectives or pharyngealized consonants. The original ejective consonants are most familiar from Ethiopic and Modern South Arabian languages. The latter, in fact, are extraordinarily conservative, retaining all of the Proto-Semitic fricative contrasts as well as the ejectives.

The morphology of the Semitic languages stands out for having a high degree of TRANSFIXATION. A transfix is a type of affix that interleaves with its base. This will be seen in depth in the sketch of Arabic at the end of this chapter (§5.6), but here is a quick preview. In Arabic, the root meaning 'write' can be thought of as the bare consonants [ktb]; in Semitic languages, most roots consist of three consonants. To form a word meaning 'writer', one adds the transfix [a:i-], which interleaves with the root to form [ka:tiḅ].

Semitic languages have a masculine-feminine gender system, similar to that of Russian (§3.5.4.1) but with no neuter. Akkadian had SOV word order, but the West Semitic languages, at least in their older, classical, forms, are VSO.

5.1.2 BERBER

Berber languages (Table 5.4) were the first Afro-Asiatic languages spoken in a large area of northern Africa west of the Nile. Their spotty distribution on today's map is due in large part to displacement by speakers of Arabic, an Afro-Asiatic language of a different branch. As often happens when languages are displaced, the original Berber languages remained strongest in less accessible locations: mostly mountain and desert areas.

TABLE 5.4

Some Berber Languages

Name	Size	Location
Northern		
· Tachelhit (SHI)	7	Morocco
· Tamazight (TzM)	7	Morocco
· Tarifit (RIF)	7	Morocco
· Kabyle (KAB)	7	Algeria
Tuareg (TMH)		
· Tawallammat Tamajaq (TTQ)	6	Niger
· Tayart Tamajeq (THZ)	6	Niger
· Timbuktu Tamasheq (TAQ)	6	Mali
· Tahaggart Tamahaq (THV)	5	Algeria
Eastern		
· Awjilah (AUJ)	4	Libya
· Sawknah (SWN)	1	Libya
Zenaga (ZEN)	4	Mauritania

By far the largest community of Berber speakers is in the Atlas mountain area of Morocco and Algeria. Some of the major Berber languages in this area include Tachelhit, Tamazight, Tarifit, and Taqbaylit. These names all look similar because language names in Berber are all feminine, and the marker of the feminine in Berber is the CIRCUMFIX [t]...[t] – that is, an affix that surrounds its base. Not infrequently, the names are also cited in English without affixes, e.g. Shilha, Riff, and Kabyle. These Northern Berber languages have a high degree of mutual intelligibility. A prominent diaspora of Northern Berber speakers is found outside of Africa, especially in France, as a result of the former French colonial presence. In addition, almost all speakers of Judeo-Berber (JBE) relocated from Morocco in the 1950s and 1960s, mostly to Israel.

A smaller but widely dispersed group of Berber speakers are the Tuareg, well known as the Muslim nomads of the Sahara whose men, not women, traditionally wear veils. They live mainly in Sahel regions of Niger and Mali and in the Algerian Sahara, but Tuareg communities are found from Burkina Faso to Libya. Their language is often referred to by the cover term *Tamasheq*. Tamasheq dialects are all very similar to each other. They form a distinct group within Berber, but are still mutually intelligible with Northern Berber.

Finally, smaller communities of Berber speakers live farther east, in Libya and Egypt; and there is also a small community in Mauritania. Owing in part to their relative isolation from other Berbers, these languages are a bit more distinct from the Northern Berber and Tuareg groups.

The name *Berber* is an exonym bestowed by Arabic speakers, from a word meaning ‘incomprehensible speech’, or ‘barbarians’, or perhaps both. It has

recently become popular to use *Tamazight* as a cover term for Berber language, especially by those who advocate linguistic, cultural, or political unity among Berbers.

Berber languages have pharyngealized consonants, transfixation, masculine-feminine gender, and VSO word order.

Berber languages have been written in Arabic, Latin, and Tifinagh scripts; in addition, Hebrew script is used for Judeo-Berber. Tifinagh is an alphabetic writing system that was used in Berber territory as early as the third century BC. The system was preserved in use through the millennia, mainly by Tuareg women. In recent times, using a modernized version of Tifinagh has become a popular expression of pride in Berber culture, and since 2003 Tifinagh has been the official script for Berber languages in Morocco. Here is the word *Tifinagh* in Tifinagh: ⵜⴰⴼⵉⴼⵏⴰⴳⴰ

5.1.3 CHADIC

The Chadic branch comprises some 195 languages spoken near Lake Chad in the Sahel (Table 5.5). The most widely spoken Chadic language by far is Hausa, a principal language of Nigeria and Niger with 25 million native speakers and 15 million second-language speakers, who use it as a lingua franca across a large expanse of West Africa. Hausa has ejective consonants and also a set of consonants distinguished by IMPLOSION – pulling down the larynx, as [ɓ] and [ɗ] – or alternatively by creaky voice, as [ɓ̰] and [ɗ̰]. It has two types of RHOTIC ([r]-like) consonants.

TABLE 5.5
Some Chadic languages

Name	Size	Location
West		
· Hausa (HAU)	8	Nigeria
Biu-Mandara		
· Bura (BWR)	6	Nigeria
· Margi Central (MRT)	6	Nigeria
· Huba (HBB)	6	Nigeria
· Kamwe (HIG)	6	Nigeria
· Mafa (MAF)	6	Cameroon
East		
· Nancere (NNC)	5	Chad
· Gabri (GAB)	5	Chad
· Mubi (MUB)	5	Chad
Masa		
· Marba (MPG)	6	Chad
· Masa (MCN)	6	Cameroon
· Musey (MSE)	6	Chad

In addition to the familiar trilled or tapped [r], it has a flap [ɾ], which is formed by curling the tongue back then flipping it forward across the alveolar ridge. Hausa is a tone language, with a high tone, low tone, or falling tone on each syllable. Hausa does not have much transfixal morphology: derivation is performed with suffixes and prefixes. All nouns have masculine or feminine gender, and the basic word order is SVO.

5.1.4 CUSHITIC

The Cushitic group (Table 5.6) takes its name from Cush, who was the son of Noah's son Ham. The Bible tells of the various regions of the world as being inhabited by Noah's immediate descendants, and a long-standing tradition associates the Biblical land of Cush with the region south of Egypt, which was known in other traditions as Nubia, Ethiopia, or Abyssinia.

While a few Cushitic languages are spoken in Kenya and Tanzania, most are spoken in Ethiopia and Somalia. There are about 45 Cushitic languages, but there are some substantial disagreements about what languages should be considered Cushitic. Somali is the principal and official language of Somalia, and the name Afar is likely to be familiar to you as the site of many important hominid fossil finds like Lucy: *Australopithecus afarensis*. Less familiar to Westerners is the Ethiopian language Oromo, even though it has one of the biggest speaker bases in Africa.

Cushitic languages tend to have masculine-feminine gender and SOV word order. Many do not have emphatic consonants, although Oromo has ejectives. Implosives and pharyngeal fricatives are common. Derivation is primarily through affixation rather than Arabic-style transfixation.

TABLE 5.6

Some Cushitic, Omotic, and Egyptian languages

Name	Size	Location
Cushitic	8	Nile, Horn
· Somali (SOM)	8	Somalia
· Hadiyya (HDY)	6	Ethiopia
· Oromo (ORM)	7	Ethiopia
· West Central Oromo (GAZ)	7	Ethiopia
· Afar (AAR)	7	Ethiopia
· Sidamo (SID)	7	Ethiopia
Omotic	7	Horn
· Wolaytta (WAL)	7	Ethiopia
Egyptian	0	Nile
· Ancient Egyptian (EGY)	0	Nile
· Coptic (COP)	0	Egypt

5.1.5 OMOTIC

Twenty-nine Omotic languages are spoken in Southwestern Ethiopia. This relatively small group has been among the most challenging to relate to its Afro-Asiatic relatives. For some time Omotic was regarded a branch of Cushitic, and a few questions have been raised about whether Omotic is an Afro-Asiatic group at all. Yet most scholars today agree that Omotic is an independent group within the Afro-Asiatic family.

5.1.6 EGYPTIAN

Records of Egyptian go back as far as 3200 BC, when early hieroglyphs were first used to occasionally label objects in artwork. This early date makes Egyptian the language with by far the longest continuous historical attestation. Sumerian writing began at approximately the same time, but that language died out as a spoken language by about 1800 BC. In contrast, Egyptian survived as a spoken language until just a few centuries ago, and even now may be used as a liturgical language among Egyptian Christians. It is conventional, however, to refer to the language as *Coptic* from about the time it came to be written in the Greek alphabet in the first centuries AD. After the Arab conquest of Egypt in the seventh century AD, the Coptic language was gradually replaced as the spoken language by Arabic.

As the language of one of the world's great empires and civilizations, Ancient Egyptian and particularly its writing continue to attract modern scholars. You are, no doubt, familiar with the appearance of Egyptian hieroglyphs, most of which depict real-world objects with a great deal of fidelity and grace. After knowledge of hieroglyphs died out, the pictorial appearance of the hieroglyphs led most would-be decipherers to assume that hieroglyphs symbolized the object they visually resembled. The discovery of the Rosetta Stone, which had an Egyptian proclamation in hieroglyphs along with a Greek translation, allowed Champollion and others to work out the true nature of Egyptian writing by 1822. They found that in any given text, only a small proportion of hieroglyphs actually represent the object they picture. To cite a few examples, the symbol ☉ typically represented the word that meant 'sun', or the sun god Ra; thus it was an iconic logogram. Other logograms were only indirectly iconic because the picture was of a homophonous word: the logogram for 'life', 𓎗, is actually a picture of something whose Egyptian name sounded like the word for 'life', perhaps a sandal strap. More often, hieroglyphs were used as phonograms. Each of the consonants of Egyptian had its own hieroglyph; additionally, many hieroglyphs could be used to represent a combination of two consonants within the same morpheme. Egyptian writing was very often redundant: a word might be spelled with both a logogram and with phonograms, and even a biconsonantal phonogram could

be accompanied by one or both of the hieroglyphs that represent its individual consonants. It was also common to add to a word a hieroglyph that acted as a semasiographic classifier. For example, the word for ‘procession’ might be spelled as in (2), where the first three symbols (reading top down in columns arranged left to right) stand for the consonants in the word, and the last tells us the word has something to do with movement.

(2) Egyptian hieroglyphs



Knowledge of the classical Egyptian consonants is incomplete, and the Egyptians did not write vowels at all. Despite this uncertainty, it does seem to be the case that they had a three-way contrast (perhaps voiced, voiceless plain, and ejective) among the stops in many places of articulation, and that the language employed transfixation in inflection and derivation. Classical Egyptian distinguished masculine and feminine gender and had a VSO word order.

5.1.7 SUMMARY

There are two features that are fairly widely distributed among the Afro-Asiatic languages. First, most of these languages had voiced, voiceless, and emphatic consonants. It should be remembered, however, that *emphatic* means ejective in some languages, pharyngealized in other languages, and sometimes implosive. Second, many Afro-Asiatic languages exhibit an unusual amount of transfixal morphology. They also tend to have a masculine-feminine gender system. The most widespread word order in Afro-Asiatic is VSO, which is notable because it is less common crosslinguistically than subject-first word orders. However, several branches and individual languages within Afro-Asiatic do have subject-first word orders.

An earlier name for the Afro-Asiatic family was *Hamito-Semitic*. Like *Semitic*, the term *Hamitic* was derived from the name of one of the sons of Noah, Ham. The Bible portrayed Ham as the progenitor of many African peoples; consequently, in the Middle Ages, Africa became considered the land of Ham. From one perspective, it made sense to refer to a language family that extended from the land of Ham to the land of Shem – that is, from Africa to the Asian near east – as *Hamito-Semitic*.

At the same time, the use of the word *Hamitic* had some unfortunate overtones. In many racist and slave-holding societies, the name *Ham* was heard primarily in the phrase “the curse of Ham”, a doctrine that held that black Africans were cursed by God to the station of slavery. In a subsequent racist development, certain nineteenth-century European scholars repurposed the term *Hamitic* to

refer exclusively to northern Africans such as Berbers and Egyptians, who looked more like Europeans than did other Africans, and so were considered superior. Well into the 1940s, most Africanists accepted language classifications that rested as often as not on racial criteria. Languages spoken by people with a more Mediterranean appearance, or by people who herded cattle (considered a typically Eurasian lifestyle), were more likely to be classed as Hamitic languages, and therefore related to the Semitic languages.

In comparison to these problems, the third issue with the term *Hamito-Semitic* seems quite benign. Linguists in the past considered the Hamitic languages to be one clade, related to but separate from Semitic. This assumption of a Hamitic branch of languages was based more on geographic or racial grounds than on cladistic theory. Whatever the actual intent of the term *Hamito-Semitic* – there is, after all, no rule that says that names of language families have to consist of the names of valid subgroups – its use tended to reinforce the idea that there was a Hamitic group of languages. In the 1950s, Greenberg carefully studied the linguistic correspondences among the so-called Hamitic languages and concluded that Hamitic's subgroups were no more closely related to one another than each subgroup was to Semitic itself. Greenberg accordingly proposed that the peoples previously classed together as Hamitic instead all spoke languages in branches that were sisters to each other and to Semitic. Greenberg insisted that racial characteristics had no bearing on how closely two languages were related. Over time, people migrate. When this happens, they sometimes abandon their native tongue in favor of the established language in the new region. That makes racial characteristics a highly unreliable indicator of the origins of a language. Likewise, lifestyles such as cattle raising change over time, so this was a really poor criterion for determining language ancestry.

After removing several languages that had been added to Hamitic on spurious grounds, Greenberg proposed renaming the family *Afro-Asiatic*, a term that had been used previously but not very widely. The new name avoids the suggestion that Hamitic is a valid grouping of languages, and it also avoids association with the racist uses to which Ham's name has been put. After some debate, most scholars came around to Greenberg's analysis.

5.2 Nilo-Saharan

The Nilo-Saharan languages (Table 5.7) are, very broadly speaking, situated between the Afro-Asiatic languages to the north and the Niger-Congo languages to the south. In comparison to those two families, Nilo-Saharan is small, with some 205 languages spoken by somewhere between 20 and 30 million people. The term *Nilo-Saharan* reflects two principal locations, the Nile region and the Sahara, and the languages in this family extend as far east as Ethiopia, as far west as Mali, and as far south as Tanzania (see the languages labeled with an initial “2” in Figure 5.1).

TABLE 5.7
Some Nilo-Saharan Languages

Name	Size	Location
Eastern Sudanic		
· Nilotic		
· · Maasai (MAS)	7	Kenya
· · Dinka (DIN)	7	South Sudan
· · Dholuo (LUO)	7	Kenya
· · Nuer (NUS)	6	South Sudan
· Nubian		
· · Old Nubian (ONW)	0	Egypt
· · Kenzi (XNZ)	6	Egypt
· · Nobiin (FIA)	6	Egypt
Central Sudanic		
· Mādi (MHI)	6	Uganda
· Moru (MGD)	5	South Sudan
· Efe (EFE)	5	Congo
· Lese (LES)	5	Congo (Republic)
· Asoa (ASV)	5	Congo (DRC)
· Mangbetu (MDJ)	6	Congo (DRC)
Saharan		
· Kanuri (KAU)	7	Nigeria
Berta (WTI)	5	Ethiopia
Songhai (SON)	7	Mali
Masalit (MLS)	6	Sudan
Fur (FVR)	6	Sudan
Kunama (KUN)	6	Eritrea
Komuz		
· Gumuz (GUK)	6	Ethiopia
Maban		
· Maba (MDE)	6	Chad

South Sudan might be thought of as the center of gravity, in that most of its indigenous languages are Nilo-Saharan.

Greenberg devised the term *Nilo-Saharan* in his highly influential African language classification (1963). Earlier experts had grouped these languages into well over a dozen separate families, often classifying many of them as Afro-Asiatic or Niger-Congo based on linguistically irrelevant criteria such as race or traditional livelihoods. In fact, the debate still goes on over subgroupings of Nilo-Saharan and over whether individual languages should be included in the family at all. But the scholarly consensus is that Nilo-Saharan itself is a valid cladistic grouping, although there is as yet no single commonly accepted reconstruction or set of sound correspondences (Blench 2000).

Nilo-Saharan has two large branches, Eastern Sudanic and Central Sudanic, and several smaller ones, including Saharan, Songhai, and Fur.

5.2.1 EASTERN SUDANIC

About half of the Nilo-Saharan languages belong to the Eastern Sudanic group. *Sudan* in this context refers to a broad savannah region that stretches from Mali to Ethiopia; the eastern Sudan therefore corresponds to the modern countries of Sudan and South Sudan.

This group includes the Nubian subgroup, spoken in Egypt and Sudan. Old Nubian is known from various written materials, mostly Christian texts written in a variant of the Coptic script dating as far back as 795 AD. Old Nubian is the ancestor of modern Nobiin.

The Nilotic subgroup of Eastern Sudanic, named after the Nile Valley, includes over 60 languages. Nilotic-speaking peoples include the majority of the population of South Sudan, where language groups like Dinka and Nuer are dominant. In Uganda, the largest number of Nilotic-speaking people speak languages from the Luo group, which extends into Kenya. The most prominent variety is Dholuo, which is the language that the father of US President Barack Obama spoke. Dholuo, along with many other languages of Africa, has ADVANCED TONGUE ROOT VOWEL HARMONY (Owino 2003). Most of its vowel phonemes come in lexically contrastive pairs, in which one vowel is pronounced with a pharynx that has been widened. The pharynx widening is typically effected by moving the root of the tongue forward, hence the name *advanced tongue root*, or ATR. Dholuo has four ATR vowels – [i̠], [u̠], [e̠], and [o̠] – and four non-ATR vowels – [i], [u], [e], [o] – and one neutral vowel, [a]. In general, all the vowels in a word have to be either ATR or non-ATR; a word cannot mix both types. The sole exception is the low vowel [a]: it may occur with either type of vowel. When affixes are added to a word, adjustments must often be made to ensure the vowel harmony. For example, the infinitive ending is [-o̠] when attached to roots that have ATR vowels, such as [kɛ̠l] ‘bring’ or [gɔ̠r] ‘nail’, but takes the non-ATR form [-o] when attached to non-ATR roots such as [kɛl] ‘scatter’ or [gɔr] ‘trim’.

Languages of the Nilotic branch, which originated in the general area of the country of Sudan, today extend to Kenya and northern Tanzania, due to southward migrations by cattle herders. Most of Kenya’s 69 languages today are Nilotic, although cumulatively they still lag behind the speaker base of Swahili, a language from the Niger-Congo family. Some present-day peoples continue the pastoral occupation of their ancestors. Probably the best known among these are the Maasai, who were among the wave of immigrants entering Kenya about five centuries ago.

5.2.2 CENTRAL SUDANIC

The Central Sudanic languages, spoken by maybe 6 million people, cover a large area, including eastern parts of the Democratic Republic of the Congo

and the entire Central African Republic. Other Central Sudanic languages are also spoken in the countries to the north, including Chad, Sudan, and Uganda. Languages of this group include Ma'di and the languages of several of the best-known pygmy groups of Congo and the agriculturalists associated with them, such as Efe and Lese, and Asoa and Mangbetu. Mangbetu is notable for having a contrast between voiced, voiceless, and prenasalized bilabial trills – [b], [p̥], and [ᵐb] (Demolin 2002).

5.2.3 SAHARAN

The Saharan group has just nine languages, five of them members of the Kanuri-Kanembu cluster spoken in the region around Lake Chad. Central Kanuri (KNC) is the dominant member of this cluster, spoken mainly in northeastern Nigeria by well over 3 million people. Kanembu (KBL), in Chad, has close to 500,000 speakers, and Manga Kanuri (KBY), spoken in Niger, about 300,000. Today's Kanuri-Kanembu region was once part of the historic Bornu Empire, which grew out of the Kanem Empire established sometime around 700 BC. Even today Kanuri is called *Bornu* by some, and Borno State in Nigeria is named after it.

5.2.4 OTHER NILO-SAHARAN LANGUAGES

Fur is the name of the largest ethnic group in Darfur, as indicated by the name of that region: [da:r fu:r] is Arabic for 'home of the Fur'. Another Nilo-Saharan people of Darfur and neighboring Chad is the Masalit.

Songhai was the language of the Songhai Empire, which once covered a large expanse of West Africa. Starting from the city of Gao in Mali around 1000 BC, Songhai rulers expanded their domain over several centuries to cover parts of present-day Niger and Nigeria in the east, all the way to a coastal region of present-day Senegal in the west. Today's Songhai is a dialect cluster with about a million speakers in Mali, Burkina Faso, Benin, Niger, and Algeria. This group is fairly distant in miles from the other Nilo-Saharan languages, and as the history of the Songhai Empire shows, this has been the case for at least a millennium. Separation from the rest of Nilo-Saharan over such a long period has led to many independent linguistic developments, so much so that some scholars even question whether it is Nilo-Saharan at all, while others accepting its Nilo-Saharan origin find evidence of mixture with other languages of the area that are not related to Nilo-Saharan genetically, such as Berber (Afro-Asiatic, §5.1) and Mande (Niger-Congo, §5.3).

5.2.5 SUMMARY

There are few broad generalizations that can be made about the vocabulary or structure of Nilo-Saharan languages. A particularly widespread grammatical feature is the fact that objects that are often found in groups tend to be named by

words that are COLLECTIVE (referring to a group of things) or plural in their most basic form; singulars are formed from them by inflectional processes (Dimmendaal 2000). In many other respects, Nilo-Saharan languages vary significantly among themselves, such that it is difficult to work out common vocabulary or grammatical features. For example, all the common word orders are attested (SVO, SOV, and VSO), as are a wide range of morphological types (analytic, agglutinative, and fusional). Many of the shared properties are common in Africa and some may well have been borrowed from neighbors. For example, most are tone languages, and implosive consonants are widespread.

5.3 Niger-Congo

Niger-Congo has over 1,500 languages, more than any other language family in Africa. The Niger-Congo family is named after the Niger and Congo Rivers, but even the vast spread suggested by the distance between those two locations fails to do justice to the extent of the Niger-Congo languages, which range from the Central African Republic and Sudan in the north to the southern tip of South Africa (see the languages that have a prefixed “1” in Figure 5.2). Below the areas occupied by speakers of the Afro-Asiatic and Nilo-Saharan languages, the Niger-Congo family accounts for the vast majority of speakers of indigenous African languages, perhaps as many as 400 million. The grouping of the languages in Table 5.8 is based on the analysis presented in Williamson & Blench (2000), in highly abridged form.

5.3.1 KORDOFANIAN

Kordofanian is a group of languages spoken in the Nuba Hills in the Kordofan region of the Sudan – far to the north of the rest of the Niger-Congo languages. They are sometimes thought of as a clade, but current thinking is that the group might constitute two or three separate Niger-Congo clades.

5.3.2 MANDE

The Mandé languages lack any vestige of the gender system that is so characteristic of Niger-Congo (§5.3.7). While some linguists take this as evidence that this clade is not related to the other Niger-Congo languages, the more widely accepted explanation is that the lack of gender is a retention of the Proto-Niger-Congo state of affairs. Under this theory, other Niger-Congo languages developed their gender system after cladogenesis had separated them from Mandé.

The 73 Mandé languages are spoken in West Africa, mainly in Mali, Guinea, Sierra Leone, Liberia, and Côte d’Ivoire, with pockets in neighboring Burkina Faso, Senegal, and the Gambia.

TABLE 5.8

Some Northwestern Niger-Congo languages

Name	Size	Location
Kordofanian		
· Tegali (RAS)	5	Sudan
· Koalib (KIB)	5	Sudan
· Katla (KCR)	5	Sudan
Mande-Atlantic-Congo		
· Mande		
· · Central Mande		
· · · Maninka (EMK)	7	Guinea
· · · Bambara (BAM)	7	Mali
· · Southwest Mande		
· · · Mende (MEN)	7	Sierra Leone
· · · Kpelle (XPE)	6	Liberia
· Atlantic		
· · Fula (FUL)	8	Guinea
· · · Pulaar (FUC)	7	Senegal
· · · Pular (FUF)	7	Guinea
· · Wolof (WOL)	7	Senegal
· Ijo-Congo		
· · Ijoid		
· · · Izon (IJC)	7	Nigeria
· · · Defaka (AFN)	3	Nigeria
· · Dogon-Congo		
· · · Dogon (DTS)	6	Mali
· · · Volta-Congo		
· · · · North Volta-Congo		
· · · · · Kru		
· · · · · Tepo Krumen (TED)	5	Côte d'Ivoire
· · · · · Pye Krumen (PYE)	5	Côte d'Ivoire
· · · · · Gur		
· · · · · Mooré (MOS)	7	Burkina Faso
· · · · · Adamawa-Ubangi		
· · · · · Gbaya (GSO)	6	Central African Rep.
· · · · · Ngbaka (NGA)	7	Congo (DRC)
· · · · · South Volta-Congo		
· · · · · Kwa		
· · · · · Akan (AKA)	7	Ghana
· · · · · Baoulé (BCI)	7	Côte d'Ivoire
· · · · · Benue-Congo	(see Table 5.9)	

The N'Ko alphabet is a recent script invented in 1949 by Solomana Kante because he felt the Mande languages should have an indigenous writing system to promote literacy as well as cultural pride and unity. The script is reminiscent of Arabic in that it is written from right to left, and the letters are all connected to each other (cf. §5.6). Unlike Arabic, however, which is an abjad (§2.1.3.3), N'ko has full letters for each of the vowels. For example, the name of the script ɔ̄ɓɔ̄ consists of ɔ̄ [ɔ̄], ɓ [k], and ɔ̄ [o].

5.3.3 ATLANTIC

The Atlantic clade is best known from Wolof and Fula. The former functions as a lingua franca in Senegal. Fula is spoken in a wide area in the North Atlantic countries from Senegal to Sierra Leone, and is perhaps the dominant indigenous language in Guinea. Thanks to the traditional livelihood of its speakers as pastoral nomads and traders, Fula is also found in many scattered localities as far east as Sudan.

Some of these languages have initial consonant apophony, or mutation (recall a similar phenomenon in the modern Celtic languages like Irish, §3.1.1). For example, in Fula, the morpheme for 'dog' has the following allomorphs depending on the suffix that follows the stem (Gregersen 1977: 86):

- (3) a. rawa:-ⁿdu 'dog'
 b. dawa:-di 'dogs'
 c. ⁿdawa-kon 'small dogs'

This example shows that the initial consonant of the morpheme meaning 'dog' alternates between [r], [d], and [ⁿd]. In Fula, there are other such apophonic series of consonants: [b] ~ [w] ~ [ⁿb], [k] ~ [h] ~ [k], and so on. Since suffixes that end in a nasal consonant (e.g. [-kon]) all require the prenasalized consonant as the initial of the stem, this consonant apophony may have arisen as a kind of assimilatory process that took place at the time when the morphemes that are now suffixes occurred before the stem morpheme.

5.3.4 KWA

Akan, the major language of Ghana, is the best studied language of the Kwa group. Its three major dialects are Asante (also spelled Ashanti), Akuapem, and Fante; the first two are often referred to collectively as Twi. The name of Anansi, the arachnid trickster of West African and Caribbean folklore, is the Asante word for 'spider'.

5.3.5 VOLTA-NIGER

The Volta-Niger clade (Table 5.9) comprises languages that have been variously classified in the past as Kwa (§5.3.4) or East Benue-Congo (§5.3.6)

TABLE 5.9
Some Benue-Congo languages

Name	Size	Location
Volta-Niger		
· Gbe		
· · Ewe (EWE)	7	Ghana
· · Fon (FON)	7	Benin
· YEAI		
· · · Yoruba (YOR)	8	Nigeria
· · Edo (BIN)	7	Nigeria
· · · Igbo (IBO)	8	Nigeria
· NOI		
· · · Nupe (NUP)	6	Nigeria
East Benue-Congo		
· Platoid		
· · Katab (KCG)	6	Nigeria
· Bantoid-Cross		
· · Cross River		
· · · Efik (EFI)	7	Nigeria
· · Bantoid		
· · · Tiv (TIV)	7	Nigeria
· · · Bantu		
· · · · Haya (HAY)	7	Tanzania
· · · · Shambala (KSB)	6	Tanzania
· · · · Swahili (SWA)	7	Kenya
· · · · Shona (SNA)	8	Zimbabwe
· · · · Kinyarwanda (KIN)	7	Rwanda
· · · · Rundi (RUN)	7	Burundi
· · · · Zulu (ZUL)	8	South Africa
· · · · Xhosa (XHO)	7	South Africa
· · · · Gikuyu (KIK)	7	Kenya
· · · · Tswana (TSN)	7	South Africa
· · · · Kongo (KON)	7	Congo (DRC)
· · · · Sotho (SOT)	7	South Africa
· · · · Northern Sotho (NSO)	7	South Africa
· · · · Ganda (LUG)	7	Uganda
· · · · Swazi (SSW)	7	South Africa

languages. In the former group are the Gbe languages, which include Fon in Benin and Ewe, which is spoken in Ghana and Togo. Ewe is often adduced as one of the few languages that phonemically contrast labiodental fricatives with bilabial fricatives, such as [v] versus [β]; the latter is the sound that is spelled

as ⟨w⟩ in the name of the language. Ewe also exhibits logophoricity. Compare these sentences:

(4) Ewe logophoricity

a. kofi be jè- \overline{dzo}

Kofi say PRO.LOG-leave

‘Kofi said that he (Kofi) left.’

b. Kofi be e- \overline{dzo}

Kofi say PRO-leave

‘Kofi said that he or she (not Kofi) left.’ (both from Sells 1987: 448)

The special logophoric pronoun (LOG) in (4a) shows that the person who left is the same person who did the saying. Uncovering logophoricity in West Africa led to the study of related phenomena in distant languages like Japanese and Finnish.

Gbe languages have very simple verb morphology, often relying on serial verb constructions reminiscent of those we encountered in Chinese (§4.10.3.8):

(5) Fon serial verbs

kòkú sɔ̀ àsɔ̀n ɔ̀ ná àsíbá

Koku take crab DEF give Asiba

‘Koku gave the crab to Asiba.’ (Lefebvre & Brousseau 2002: 446)

As this example shows, some Niger-Congo languages lean toward the analytic end of the morphological spectrum. Others are highly agglutinative. See the sketch of Swahili (§5.7) for an example of the latter.

The Volta-Niger languages that are sometimes considered to be East Benue-Congo include some widely spoken languages of Nigeria: Yoruba, Igbo, Edo, and Nupe.

5.3.6 EAST BENUE-CONGO

East Benue-Congo is by far the largest group in Niger-Congo, by any measure: it has perhaps half a billion speakers of some 900 languages, extending over most of sub-Saharan Africa. The name comes from two large rivers, the Benue, running from Cameroon to neighboring Nigeria, and the much larger Congo, which begins in Zambia and passes through a number of East and Central African countries on its way to the Atlantic Ocean. The language group has several subbranches, of which the Cross River clade is one of the most prominent, with Efik having the most speakers. The Bantu subbranch contains a very large number of languages and occupies a larger territory than all the other subbranches of Niger-Congo put together. It is believed that Bantu languages originated in Nigeria and Cameroon

and gradually spread as far as South Africa, starting about 2,000 years ago. Bantu is a subgroup of a larger group, Bantoid, which includes 169 neighboring languages still spoken in Cameroon and Nigeria.

The Bantu languages of South Africa, such as Xhosa and Zulu, have been influenced by the Khoisan languages of the Kalahari desert (§5.4). For example, they have borrowed CLICKS from those languages – consonants produced by a velaric airstream. Swahili is used as an official language in many countries of East Africa and as a lingua franca in adjacent regions. See the Sketch of Swahili (§5.7) for more details about this language. Bantu languages are marked by extensive use of affixes marking gender, verb tense and aspect, agreement markers on verbs, and a variety of derivational categories. Bantu languages are among those linguists most often use to illustrate agglutinative languages.

5.3.7 SUMMARY

The great majority of Niger-Congo languages are tonal. Exceptions are found in the Atlantic branch; another major exception is Swahili. The basic word order is most often SVO, but the Ijoid languages and Mande are SOV.

Many languages in West Africa share an areal feature that is not very common elsewhere in the world: DOUBLY ARTICULATED stops. Hundreds of African languages have phonemes in which a velar and bilabial closure overlap: [g̠b], [kp̠], and, more rarely, [ŋ̠m]. That is, the lips close before the velar closure is released. An example is Yoruba [g̠bà] ‘receive’.

In Niger-Congo languages, simple nouns are combinations of an affix plus noun stem, although in some languages, such as Igbo, Efik, and Yoruba, this structure is very poorly represented. Most subbranches of Niger-Congo have a gender system. The number of genders is much larger than the three genders found in Indo-European, and they are not connected with sex or biological gender, unlike the masculine and feminine genders of Indo-European and Afro-Asiatic. These gender systems are especially prominent in the Bantu languages. In Swahili, for example, there are at least seven separate genders (§5.7). As in Indo-European languages, adjectives, demonstratives, and other types of noun modifiers have to agree in gender and number with the noun they modify.

Tone in Niger-Congo and in other African language families differs in several respects from tone in Mandarin (§4.10.2.3) and other East Asian languages. At first glance, the African systems appear simpler, because many languages have just two basic tones, high and low, and those basic tones are normally heard as level rather than rising or falling. This simplicity is reflected in choice of IPA transcription, where diacritics like [ˈ] high and [ˌ] low are usually preferred to the more complicated tone letters used for Asian languages, where many contour tones are found. Still, a number of languages have three or four tone levels. Furthermore, and more spectacularly, even in languages with just

two tone levels, tones can change in a surprising number of ways, depending on where they appear in a word or phrase. For example, in Haya (5.3.6), a high tone automatically changes to low if it happens to be at the very beginning of an utterance, while a noun whose last two syllables respectively have high tone and low tone will shift that pattern to low tone followed by high tone at the end of an utterance. When all the relevant tonal changes in Haya are taken into account, a single three-syllable word can have as many as eight different tone patterns, depending on what comes before and after.

Another common characteristic of tone in Niger-Congo and some other African languages is *DOWNSTEP*, a lowering of the tonal pitch by a more or less fixed degree. Pitch lowering sets a new, lower ceiling for subsequent tones in the phrase. If a high tone is downstepped after another high tone, the second will sound lower in pitch than the preceding while generally remaining distinct from a low tone. The effects of downstep are cumulative: if a phrase includes a succession of downsteps, each one lowers the pitch from the previous level.

In some languages, it is easy to trace the occurrence of downstep back to a case in which a low tone has been deleted by a general rule. Such is the case in the Nigerian language Tiv (§5.3.6). Compare these two sentences:

- (6) a. Tiv
 kásév ⁿbâ
 women COP
 ‘There are women.’
- b. Kásév ⁿbá⁺ gá
 women COP NEG
 ‘There aren’t any women.’ (both from Arnott 1964)

Some Tiv words, like [ⁿbâ], have two tones, high followed by low, on a single syllable. We hear both tones when [ⁿbâ] appears at the end of a sentence, as in (6a), where the tonal sequence high–low is heard as a fall in pitch. (It is no coincidence that the IPA symbol for falling tone, [^{˥˩}], is literally a high-tone symbol joined to a following low-tone symbol.) But elsewhere in an utterance, Tiv allows only one tone per syllable. Thus in (6b), the second tone of [ⁿbâ] is not pronounced. Nor is it totally obliterated. The low tone component of [ⁿbâ] causes a downstep of the pitch of the subsequent high tones, which can be indicated in IPA by placing [⁺] where the lowering in pitch begins. But because the second, low-tone component of [ⁿbâ] cannot be pronounced on its home syllable in the middle of a phrase, the sequence sounds like a high tone on [ⁿbâ], that is, [ⁿbá], followed by a downstepped high on [gá].

Not all cases of downstep are this easy to trace. In some languages, downstep is *DEMARCATIVE*: it helps indicate where one word stops and the next one begins. Shambala (§5.3.6) places a downstep between any two words in a phrase, as long as the first ends in a high tone and the second begins with one.

(7) Shambala

ázàkómá † nóká
 killed DEMARC snake

'He killed a snake.' (Odden 1982)

Understanding tone in Niger-Congo languages has shed light on the behavior of tone and accent in languages worldwide. Cases like the Tiv examples (6) have shown, for example, the explanatory power of analyzing a falling contour tone as a sequence of two simple level tones. The analysis of downstep in African tone languages has provided a model for intonational phenomena in typologically distinct languages from distant corners of the globe, including Japanese and English.

5.4 Khoisan area

The Khoisan languages (Table 5.10) are found primarily in the Kalahari desert region of southwestern Africa (the languages that have a prefixed “2” or “3” in Figure 5.2), but two language isolates in Tanzania have also been included in this grouping (SAD and HTS in the map). Originally these languages occupied a much larger territory, but they were then pushed into the less hospitable desert areas by Bantu tribes from the north and white settlers from the south. Now only about 300,000 people speak a Khoisan language, most of these being speakers of Nama.

Greenberg (1963) proposed that the Khoisan languages form a language family, but the current consensus is that there are insufficient sound correspondences

TABLE 5.10
Some languages of the Khoisan area

Name	Size	Location
!Kung		
· Kung-Ekoka (KNW)	4	Namibia
· †Kx'au 'ein (AUE)	4	Namibia
· Ju 'hoan (KTZ)	4	Namibia
Khoe		
· Nama (NAQ)	6	Namibia
· Hai om (HGM)	5	Namibia
· Khwe (XUU)	4	Namibia
Taa-!Kwi		
· †Hua (HUC)	3	Botswana
Sandawe (SAD)	5	Tanzania
Hadza (HTS)	3	Tanzania

to come to that conclusion. The more conservative assessment is that these languages should be grouped into several different families.

The !Kung language family (the languages that have a prefixed “2” in Figure 5.2) has also been called *Northern Khoisan*, but that term is now avoided lest it support the idea that Khoisan is a language family. The apparent exclamation mark is the symbol for a type of click (Table 5.11).

The Khoe family (the languages that have a prefixed “3” in Figure 5.2), earlier called *Central Khoisan*, is represented by Nama, which is by far the most widely spoken of the Khoisan languages. It also goes by the name of Khoekhoe. An earlier name was *Hottentot*, a word fabricated by colonists to suggest that the language was incomprehensible babble; it is now considered derogatory. The Khoe languages include an impressive number of clicks, but they are eclipsed by the complexity of the !Kung phonology.

The Taa-!Kwi family is approximately the erstwhile *Southern Khoisan*. It is now named after its two branches; the simpler name *Tuu* has also been proposed.

The main reason for originally classifying all the Khoisan languages into a single family was that all of them have clicks. Clicks are typologically rare, not being attested in any ordinary spoken language outside of Africa. By the same reasoning, it was once thought that the presence of clicks in Sandawe and Hadza, two language isolates spoken in Tanzania, meant that those languages were also genetically Khoisan. Today it is generally conceded that a single typological feature is not enough to show that languages are related. Indeed, there is no doubt that clicks can be borrowed by unrelated languages, as has happened in some neighboring Bantu languages, including Xhosa and Zulu, and in Dahalo DAL, a Cushitic language spoken in Kenya, which has clicks in a few dozen words.

Table 5.11 shows the symbols used in Khoisan and Bantu languages for click sounds. The IPA uses the Khoisan symbols. In addition to these basic types of clicks, Khoisan languages also have many phonemically contrastive variants, including those that are voiced [ɓ], aspirated [ɓʰ], nasal [ɓ̃], glottalized [ɓ̚], and affricated [ɓ̥χ]. This proliferation of clicks contributes to very large consonant inventories in Khoisan languages, rivaling those of the West Caucasian languages

TABLE 5.11

Symbols used to transcribe clicks

Kind	Khoisan orthography	Bantu orthography
Bilabial	⊙	
Dental		c
Retroflex	!	q
Palatal	‡	
Lateral		x

(§3.3.1). The vowel inventories are sometimes complex as well. Many languages have a distinction of nasalized versus oral vowels and also pharyngealized vowels.

The most common basic word order of Khoisan languages is said to be SVO, but OSV is also reported for some languages. Some languages have grammatical gender based on sex. Verb morphology is complex, and sometimes totally different lexical items are used to express the same action depending on whether it is being performed by a single person or by more than one person.

5.5 Other languages in Africa

5.5.1 AUSTRONESIAN

Yet another language family is found in Madagascar, a large island off the southeastern coast of Africa, which was settled by a group who migrated from the Malay archipelago more than 1,500 years ago. Malagasy (MLG), the national language of Madagascar's nearly 20 million inhabitants, descended from the Malayo-Polynesian language spoken by the island's earliest settlers. Its nearest relatives are some western Austronesian languages spoken on Borneo. We'll read more about that language family in §6.1. The basic word order in Malagasy is VOS. Malagasy has borrowed vocabulary items from various Bantu languages such as Swahili, from Arabic, and from such European languages as French and English, but its Austronesian genetic affinities are not in doubt.

5.5.2 INDO-EUROPEAN

Finally, the languages of European colonizers still play important roles in many African countries today, often sharing official status with African languages. For that reason several languages of the Indo-European family are also widely represented on the continent. These are Afrikaans, English, French, and Portuguese.

English and French are still very influential in many former British and French colonies in Africa, especially in those African countries in which a large number of very different native languages are spoken and there are not yet clear national languages emerging. A number of noteworthy African writers and poets write primarily in English or French.

The case of Afrikaans (§3.1.2) is special in that it is not merely an imported colonial language but something that grew and developed in South Africa from the language of the Dutch settlers, who started arriving there in the latter half of the seventeenth century. It is now considered to be a language separate from Dutch. Afrikaans is spoken in South Africa, Namibia, and surrounding African countries. It is widely used as a second language throughout the region; of 20 million speakers, perhaps only some 7 million are native speakers. Some aspects of its grammar are simpler than in Dutch. It contains many

loanwords from Malay, Portuguese, other European languages, and various African languages.

Because until relatively recently Africa had been under the political control and strong cultural influence of the various European powers, there have arisen in Africa many contact languages, pidgins, and creoles, several of which will be discussed in Chapter 8.

5.6 Sketch of Modern Standard Arabic

5.6.1 GENETIC RELATIONSHIP AND GENERAL INFORMATION

Arabic belongs to the Semitic branch of the Afro-Asiatic family of languages, more specifically, to its Central subbranch. The oldest Arabic inscriptions go back to the fourth century AD, but inscriptions are not numerous until the Islamic period, which began in the seventh century.

Modern Arabic has a large number of dialects, many of which are not mutually intelligible. The greatest differences are found between Maltese, which is often considered a separate language, and the rest. Aside from Maltese, regional dialects form two main groups: (1) Eastern Arabic, which includes dialects spoken in Mesopotamia, Syria, Lebanon, Israel, the Arabian peninsula, Egypt, Sudan, and Zanzibar in Tanzania; and (2) Western Arabic, which includes dialects spoken in North Africa: Libya, Morocco, Algeria, Tunisia, and Mauritania. In addition, in each region there is also a significant dichotomy between the dialects of the city dwellers and those of rural or desert dwellers, such as the Bedouins. The dialect with the largest number of speakers is Egyptian Arabic (ARZ), with 54 million speakers.

A formal, internationally standardized variety of Arabic called *fuṣḥā* فصحى [fusˤħa:] coexists along with the vernacular varieties of Arabic. *Fuṣḥā* is the language of the Koran (قرآن [qurʔa:n]), the holy scripture of Islam, which was recorded in the seventh century AD. This language is used even now for communication between speakers of different Arabic dialects, for formal speeches, formal documents, serious literature, and so forth, whereas the local dialect is used primarily for ordinary oral communication and for literature that is considered less serious or that strives to portray everyday life. This situation, in which there are two very different variants of the same language used in the same community for separate purposes, is called *DIGLOSSIA*. In addition, linguists distinguish two forms of *fuṣḥā*: Classical Arabic (ARB), the language written and spoken in Koranic times, and Modern Standard Arabic, the language as it is actually used today. Although Modern Standard Arabic is based on classical norms, many differences are apparent, especially in vocabulary and pronunciation.

Muslims are expected to learn Classical Arabic in order to be able to read and understand their holy scripture in its original form. All prayers are also said in Classical Arabic. This means that in non-Arab Muslim countries as well, Classical

Arabic exerts a great influence on the local languages, for instance, on Persian in Iran and on Turkish in Turkey. In turn, the Arabic language is itself influenced by the surrounding languages, especially Persian and Turkish, and in recent times it has also borrowed words from European languages, especially French, English, and Italian. The regional dialects of colloquial Arabic differ in the extent to which they have borrowed from different European languages. Thus, for example, North Levantine Arabic (APC) (Syria and Lebanon), Tunisian Arabic (AEB), Moroccan Arabic (ARY), and Algerian Arabic (ARQ) have borrowed more from French; Libyan Arabic (AYL) more from Italian; and Egyptian Arabic and Sudanese Arabic (APD) more from English.

The English language has quite a few Arabic loanwords. Many of these loanwords entered English via Spanish, which itself borrowed many words from Arabic during the centuries when parts of the Iberian peninsula were Muslim territory ruled primarily by Arabic speakers (711–1492). Some of the loanwords can be easily spotted because they contain the Arabic definite article [al-]:

- ☒ alcohol
- ☒ algebra
- ☒ alchemy (originally Greek)

Others do not contain the definite article:

- ☒ syrup
- ☒ sherbet (Both this word and *syrup* derive ultimately from the Arabic root [ʃrb] ‘drink.’)
- ☒ nadir
- ☒ zenith

Many astronomical terms are of Arabic origin, including names of some stars, such as Aldebaran.

5.6.2 PHONETICS AND PHONOLOGY

5.6.2.1 Consonants

There is much variation in the pronunciation of Modern Standard Arabic. Modern Standard Arabic is not a native language of any group of speakers, and therefore it tends to be influenced by the sound systems of local dialects. Table 5.12 lists the consonants in a conservative standard pronunciation.

The pharyngealized consonants (those transcribed with a “^h”) are the emphatic counterparts of the corresponding consonants. The panoply of sounds produced in the throat and in the back of the mouth is considered very characteristic of Arabic, although these features are shared to a certain extent by other Semitic and Afro-Asiatic languages.

Some common variations include pronouncing [ð^h] as [z^h]; pronouncing uvular fricatives as velars; and, especially in Egypt, pronouncing [d̥ʒ] as [g]. Although

TABLE 5.12

Consonant phonemes of Modern Standard Arabic

Stop			t, t ^s	k	q	ʔ
· Voiced	b		d, d ^s			
Affricate				ḏ		
Fricative	f	θ	s, s ^s	ʃ	χ	ħ
· Voiced		ð, ð ^s	z		ʁ	ʕ
Nasal	m		n			
Approximant			l, r	j	w	

[ʕ] is usually described as a voiced pharyngeal fricative, it is typically reinforced by some glottal activity that makes it easier to perceive; some speakers use creaky voice [ʕ̤] or turn it into a pharyngealized glottal stop [ʔ^h]. Its counterpart [ħ], on the other hand, is a true voiceless pharyngeal fricative.

It is interesting to note the gaps in the system. For example, there is no [p] corresponding to [b]; Proto-Semitic [p] turned into [f] in Arabic. There is also no [g] corresponding to [k]; Proto-Semitic [g] has turned into [ḏ] in most Arabic dialects.

5.6.2.2 Vowels

Modern Standard Arabic has a triangular three-vowel phonemic system – [a], [i], and [u] – which is doubled to six by the fact that vowel length is also phonemic. Vowels next to pharyngeal, pharyngealized, and uvular consonants have very different allophones from those next to the other consonants. The former are lower and more back. For example, the phoneme /a/ is realized as [æ] or even as [ɛ] next to plain consonants, whereas it is realized as [ɑ] in the environment of pharyngeal, pharyngealized, or uvular consonants.

5.6.3 ARABIC ORTHOGRAPHY

The Arabic script developed from the Phoenician script, as did Latin and many other scripts. In particular, it developed from Aramaic versions of the abjad, as did the modern Hebrew script (§2.2.2). As Table 5.13 illustrates, Arabic was optimized for speed of writing, with so many simplifications that in the early Arabic texts, many letters had become identical. These ambiguities were resolved by adding diacritic dots to distinguish consonants from each other. In addition, dots were added to some letters to represent consonants not found in Aramaic ([θ], [χ], [ð], [d^s], [ð^s], and [ʁ]). All of these dots are now considered mandatory, except in certain specialized forms of calligraphy.

The Arabic script consists primarily of letters for the consonants, making it an abjad, which is, in fact, an Arabic name for the Arabic script. However, the letters for [w], [j], and [ʔ] are pressed into service to represent most

TABLE 5.13

**Changes in letter shapes
between Phoenician and Arabic**

Sound	Phoenician	Aramaic	Arabic	
			Early	Dotted
ʔ	𐤀	𐤁	ا	إ
b	𐤂	𐤃	ب	بّ
t	𐤄	𐤅	ت	تّ
p > f	𐤆	𐤇	ف	فّ
q	𐤈	𐤉	ق	قّ

instances of the long vowels [u:], [i:], and [a:]. Vowels, as well as several other details of pronunciation, can also be represented by diacritic marks, but these are used only when special precision is needed, such as in religious texts or as an aid to learners.

Arabic is written from right to left. There is no contrast between block letters and connected cursive: all writing is cursive. Nor is there a distinction between upper- and lowercase, as is present in Latin and Cyrillic. But letters may have up to four different ALLOGRAPHS, alternative forms that are selected depending on whether they are joined up to the preceding letter, the following letter, both, or neither. Table 5.14 shows their two most distinctive forms: when connected to no letter (middle column) and when joined to letters on both sides. When letters are not joined to the following letter to their left, they tend to end with a flourish, which makes them more distinctive. Joining is not up to the whim of the writer. Most letters must be joined to the following letter in the same word, and the remaining letters (marked by a * in the table) can never be connected to the following letter.

In addition there are some special ligature forms. The only one considered obligatory is that a la:m followed by an ʔalif is written as ڤ.

Except for ʔalif, the formal letter names, shown in the first column of Table 5.14, are acrophonic: the first phoneme in the letter name is the consonant that the letter represents. The special letter ڤ spells [t] at the orthographic end of a word when it spells the feminine suffix. The curious representation (ڤ ha:ʔ with the dots of ڤ ta:ʔ) reflects the fact that this suffix was often pronounced [h], especially at the end of a sentence.

There is also a diacritic, called *hamza*, which represents [ʔ]. It sometimes is written on the line of text, like a letter, especially at the end of an orthographic word (8a). More often, it is attached to an ʔalif, wa:w, or ja:ʔ, following complicated rules that give some hint as to the identity of an adjacent vowel.

TABLE 5.14

The Arabic alphabet

Name	Free	Joined
ʔalif	ا	آ
ba:ʔ	ب	بـ
ta:ʔ	ت	تـ
ta:ʔ marbu:tʔa	ة	ةـ
ḥa:ʔ	ث	ثـ
dʒi:m	ج	جـ
ħa:ʔ	ح	حـ
ħa:ʔ	خ	خـ
da:l	د	*دـ
ḍa:l	ذ	*ذـ
ra:ʔ	ر	*رـ
za:j	ز	*زـ
si:n	س	سـ
ʃi:n	ش	شـ
sʔa:d	ص	صـ
dʔa:d	ض	ضـ
fʔa:ʔ	ط	طـ
ḍʔa:ʔ	ظ	ظـ
ʔain	ع	عـ
ɣain	غ	غـ
fa:ʔ	ف	فـ
qa:f	ق	قـ
ka:f	ك	كـ
la:m	ل	لـ
mi:m	م	مـ
nu:n	ن	نـ
ha:ʔ	ه	هـ
wa:w	و	*وـ
ja:ʔ	ي	يـ

*Letter never connects with the following letter, to its left.

The ʔ base letter to which the hamza attaches is silent, and ja:ʔ used this way omits its two dots.

- (8) a. نساء [nisa:ʔun] 'women'
 b. أنت [ʔanta] 'you'
 c. إنسان [ʔinsa:nun] 'a human being'
 d. مائدة [ma:ʔidatun] 'a table'

As its name suggests acrophonically, the letter ʔalif originally represented [ʔ], but there were so many exceptions that the diacritic hamza became obligatory for all phonemic glottal stops. Consequently ʔalif now has several other roles rather than its original function. In addition to serving as a support for the hamza diacritic, it is required at the beginning of a word that begins with a vowel (9a). It also marks most instances of the vowel [a:], and also the indefinite accusative ending [-an] (9b), except after ɔ (9c).

- (9) a. الكتاب [al=kita:bu] ‘the house’
 b. بيتا [baitan] ‘a house’ ACC.INDF
 c. جميلة [ḍʒami:lata] ‘beautiful’ F.ACC.INDF

In addition to spelling [w], the glide wa:w represents long [u:] and the diphthong [au]. Analogously, ja:ʔ spells [i:] and [ai]. Less commonly, ja:ʔ can spell [a:] at the end of a word, in which case the dots are omitted.

- (10) a. شديد [ʃadi:dun] ‘strong’
 b. عين [ʔainun] ‘an eye’
 c. عروس [ʔaru:sun] ‘a bride’
 d. إلى [ʔila:] ‘to’

5.6.3.1 Optional diacritics

In addition to the required diacritics, there are several optional ones, which are used only when unusual amounts of precision are required, such as in printing the Koran. Vowels are marked by diacritics written above or below the letter for the consonant they are pronounced after, here illustrated with da:l [d] (11). Long and short vowels use the same diacritics, except that there are special diacritics for long [a:] when that is not preceded by any consonant. There is even a special sign for the absence of a vowel (11e).

- (11) a. [a] دَ, [ai] دَي, [au] دَو
 b. [a:] اَ, دَا, دَي, دُ, [ʔa:] آ
 c. [i] دِ, [i:] دِي
 d. [u] دُ, [u:] دُو
 e. ∅ دُ

There is a diacritic ˘ that can be used to indicate that a consonant is long:

- (12) خَمَّارُ ʕam:a:ru ʕam˘˘ ‘wine merchant’

The indefinite suffix [-n] is written by doubling the diacritic symbol for the preceding vowel. Note the special form that the symbol for [u] takes when doubled (13b).

- (13) a. χ amr-i-n خَمْرٍ wine-GEN-INDF
 b. χ amr-u-n خَمْرٌ wine-NOM-INDF
 c. χ amr-a-n خَمْرًا wine-ACC-INDF

5.6.3.2 Use for other languages

The Arabic script has been used to write many languages besides Arabic: Urdu (Indo-European) in Pakistan, Pashto and Dari (Indo-European) in Afghanistan, Uighur (Turkic) in China, Tibetan dialects (Sino-Tibetan) spoken by Tibetan Muslims in Kashmir, Persian (Indo-European) in Iran, and so forth. Most Turkic languages used to be written in Arabic script, but in modern times most have switched to the Latin script (as in Turkey itself) or to Cyrillic-based alphabets (as in the former Soviet Union). Diacritics are added to Arabic letters in order to write sounds that are not found in Arabic.

5.6.4 MORPHOLOGY

5.6.4.1 Noun morphology

Nouns are inherently either feminine or masculine in gender. To a certain extent, gender correlates with meaning. For example, nouns referring to females, body parts that come in pairs, and most countries are feminine. Masculine nouns are unmarked for gender, whereas most, though not all, feminine nouns have the suffix [-at] *ə*. This suffix is usually pronounced simply as [-a] in pausa. In formal usage, IN PAUSA means at the end of an utterance, but in ordinary spoken Arabic, it is most common to treat virtually every word as being in pausa. Thus, in normal spoken use, the feminine suffix is usually pronounced [-a] at the end of a word.

Nouns are marked for one of three cases – nominative, accusative, and genitive – by means of suffixes (Table 5.15). The nominative case is used to mark the subject of a verb; accusative to mark the object and to derive adverbs and prepositions from nouns; and the genitive to mark nominal dependents of nouns, such as possessors, and is the case taken by objects of prepositions. In some categories, including any number other than singular, there is no distinction between accusative and genitive. That is, such nouns are either nominative or other. A general term for “other” in the sense of non-nominative is the OBLIQUE CASE (OBL).

- (14) ‘wine’
 a. χ amr-u NOM
 b. χ amr-a ACC
 c. χ amr-i GEN

TABLE 5.15

Arabic noun inflections

Number	State	NOM	ACC	GEN
SG ^a	INDF	-un	-an	-in
	DEF, HEAD ^b	-u	-a	-i
				OBL
DU	INDF, DEF	-a:ni	-aini	
	HEAD ^b	-a:	-ai	
PL.M	INDF, DEF	-u:na	-i:na	
	HEAD ^b	-u:	-i:	
PL.F	INDF	-a:tun	-artin	
	DEF, HEAD ^b	-a:tu	-arti	

^aSingular endings are also used for transfixed plurals.

^bGoverning a genitive noun.

In pausa, case endings are normally omitted, so that one would say [χamr] regardless of case.

The PROCLITIC [al] is prefixed to a noun to show that it is definite, much like the definite article *the* in English. A proclitic is like an enclitic, except that it is attached to the beginning of a word rather than its end. The [a] of the definite article is typically dropped after a vowel in the preceding word, and the [l] assimilates fully to a following coronal consonant. Exceptionally, [l] does not assimilate to [d̪ʒ], which was originally a velar consonant.

- (15) a. al=χamr-u 'the wine' DEF-wine-NOM
 b. aʃ=ʃams-a 'the sun' DEF-wine-ACC
 c. ar=rad̪ʒul-i 'the man' DEF-wine-GEN

Most nouns take an additional suffix [-n] when indefinite (INDF) (13). In pausa, this suffix, as well as the preceding case ending, is normally dropped, leaving indefinite forms like [χamr]. Those nouns that do not take the indefinite suffix [-n] have nominative [-u] and oblique [-a] when they are indefinite. Such forms are called *diptotes*, from a Greek word meaning 'two cases'.

Arabic, like many other Semitic languages, has a distinctive set of rules that apply when one noun depends on another, to express, for example, possessive relations as in English *the man's house*. The head noun ('house' in 16) takes whatever case is appropriate for its use in its parent construction; the dependent ('man') takes the genitive case. The dependent noun immediately follows the head noun; not even an adjective may intervene. The head noun takes neither the definite proclitic nor the indefinite suffix; its definiteness is taken to match that of the dependent, where definiteness is explicitly marked.

- (16) a. bait-u raḍʒul-i-n
 house-NOM man-GEN-INDF
 ‘a man’s house’, ‘a house of a man’
- b. bait-u r=raḍʒul-i
 house-NOM DEF=man-GEN
 ‘the man’s house’, ‘the house of the man’

If the head noun ends in the feminine suffix *ə*, that suffix is pronounced as [-at], even in pronunciation styles that elsewhere tend to pronounce word-final *ə* as [-a].

Nouns can be singular, dual, or plural. As is common in most languages, the singular has no special marker. The dual (DU), which indicates two objects, is always formed regularly. The dual suffixes distinguish nominative from oblique case. They do not have different forms when definite and indefinite; they never add [-n]. When dual suffixes are added to the feminine suffix [-at], the [t] in the latter is always pronounced, because it is no longer in pausa. Dual suffixes drop their final syllable when followed by a dependent genitive.

- (17) a. [bait-a:ni] house-NOM.DU
 b. [al=bait-a:ni] DEF=house-NOM.DU
 c. [bait-aini] house-DU.OBL
 d. [malik-at-a:ni] king-F-NOM.DU ‘two queens’
 e. [bait-a: l=malik-i] house-NOM.DU DEF=king-GEN ‘the two houses of the king’

Some nouns denoting men form their plural with the following suffixes, which vary with case. Like the dual, these endings drop their final suffix before dependent genitives.

- (18) a. [muʃal:im-u:na] teacher-NOM.PL
 b. [muʃal:im-i:na] teacher-OBL.PL
 c. [muʃal:im-i: l=malik-i] teacher-OBL.PL DEF=king-GEN ‘the king’s teachers’

Many other words take the suffix [-a:t] in the plural, which is followed by case endings and, for indefinite nouns, the suffix [-n]. Before [-a:t], the feminine suffix [-at] is omitted, with the result that [-a:t] can be interpreted as the plural form of [-at]. But [-a:t] is also used with many nouns that are masculine in the singular.

- (19) a. [malik-a:t-u-n] king-F.PL-NOM-INDF ‘queens’
 b. [al=malik-a:t-i] DEF=king-F.PL-OBJ ‘the queens’
 c. [al=ħajawa:n-a:t-u] DEF=animal-F.PL-NOM ‘the animals’ (SG: [al=ħajawa:n-u] M)

Very many nouns form their plurals by means of transfixes. There are many different plural transfixes, and few reliable rules for predicting which transfix is used with which noun.

- (20) Transfix [①i②a:③]
 a. [raḍʒul-u-n] ‘man’, [riḏ̣ʒa:l-u-n] ‘men’
 b. [kalb-u-n] ‘dog’, [kila:b-u-n] ‘dogs’

- (21) Transfix [①u②u③]
 a. [kita:b-u-n] ‘book’, [kutub-u-n] ‘books’
 b. [safii:n-at-u-n] ‘ship’, [sufun-u-n] ‘ships’

The circled numbers in our notation for transfixes indicate which consonants of the root align with which position in the transfix. Most Arabic roots have three consonants. Transfixation is such a pervasive method of derivation and inflection in Arabic that even the singular form of simple nouns is usually viewed as being formed by applying a transfix to a purely consonantal root. For example, in [kita:bun] ‘book’, the root is [k-t-b].

5.6.4.2 Adjective morphology

Adjectives behave very much like nouns and, from a morphological standpoint, do not really form a class separate from that of nouns. Because of agreement rules (§5.6.5.2), they can take gender, case, indefiniteness, and number suffixes, and even transfixed plurals and the definite proclitic [al=]. The most common adjective-forming suffix is [-ij:] (in pausa, [-i:]); adjectives derived by means of this suffix can also act as nouns:

- (22) a. [mis^ʕr-u] ‘Egypt’ (diptote noun)
 b. [mis^ʕr-ij:-un] ‘Egyptian’ (adjective or noun)

The comparative and superlative degrees are not distinguished morphologically. They are expressed by a diptote formed with the transfix [ʔa①②a③]:

- (23) a. [kabi:run] ‘great’
 b. [ʔakbaru] ‘greater’, ‘greatest’

5.6.4.3 Personal pronouns

The personal pronouns are listed in Table 5.16. Note that Arabic pronouns, like nouns, have a dual number. In addition, Arabic is unusual in distinguishing gender not only in the third person but also in the second person. There is no gender distinction in the dual, however, for any of the persons.

TABLE 5.16

Arabic personal pronouns, free forms

Person and gender	SG	DU	PL
1	?ana:	naħnu	naħnu
2.M	?anta	?antuma:	?antum
2.F	?anti	?antuma:	?antun:a
3.M	huwa	huma:	hum(u:)
3.F	hija	huma:	hun:a

TABLE 5.17

Arabic personal enclitic pronouns

Person and gender	SG	DU	PL
1	=i: ^a	=na:	=na:
2.M	=ka	=kuma:	=kum
2.F	=ki	=kuma:	=kun:a
3.M ^b	=hu	=huma:	=hum
3.F	=ha:	=huma:	=hun:a

^aPreceding short vowel is elided. [=ja] after a long vowel or diphthong; [=ni:] on verbs.

^b[=hu-] in all 3rd-person suffixes becomes [=hi-] after [i].

Arabic can also express pronominal references by means of enclitics (Table 5.17), which mark possession when attached to nouns. They are attached to nouns in the form the nouns take when head of a genitive construction, as summarized in Table 5.15. Thus ‘your two teachers’ is [muʔal:im-a:=ki]. Attached to prepositions or verbs, they serve as the objects of those words. Because these forms attach to words of different classes, they are enclitics rather than inflectional endings.

5.6.4.4 Verb morphology

Arabic has a very rich verb morphology involving a large number of inflectional and derivational patterns. Like other Semitic languages, it has consonantal roots to which various transfixes are added, supplying vowels and sometimes additional consonants. These transfixes allow speakers to generate an impressively large set of derived and inflected forms that would take several pages just to list. Most roots have three consonants; patterns for shorter and longer roots will not be discussed in this sketch.

There are about a dozen derivational patterns, illustrated in Table 5.18. The circled numbers stand for the respective consonants in a verb root that has three consonants. For example, if the first row is [kataba] ‘he wrote’ in the perfective, then there is potentially a derived form that has the perfective [kat:aba] (second

TABLE 5.18

Verb derivation patterns in Arabic

PFV	IPFV	Meaning
⓪a⓪V⓪a	ja⓪⓪V⓪u	VERB
⓪a⓪:a⓪a	ju⓪a⓪:i⓪u	VERB intensely
⓪a:⓪a⓪a	ju⓪a:⓪i⓪u	VERB to someone
a⓪⓪a⓪a	ju⓪⓪i⓪u	cause to VERB
ta⓪a⓪:a⓪a	jata⓪a⓪:a⓪u	VERB oneself intensely
ta⓪a:⓪a⓪a	jata⓪a:⓪a⓪u	VERB each other
in⓪a⓪a⓪a	jan⓪a⓪i⓪u	get VERBED
i⓪ta⓪a⓪a	ja⓪ta⓪i⓪u	VERB oneself
ista⓪⓪a⓪a	jasta⓪⓪i⓪u	request VERBING

Note: V means that verbs differ on which short vowel appears in this position.

row), one with the perfective [ka:taba], one with [aktaba], and so forth. Each of these derived patterns modifies the meaning of the basic verb (i.e. the verb in the first row) in a typical way, as described in the last column. Keep in mind that derivation is unpredictable: many verbs will not take all possible derived forms, and many that do take a particular form will not have the predicted meaning. For example, there is no [takat:aba]; and while [ka:taba] does mean ‘corresponded with someone’, and [aktaba] means ‘made someone write’, as the table suggests, [inkataba] means ‘subscribed’, not ‘got written’. Each of the patterns can take all the inflections that will be discussed presently, though for lack of space we will focus on the basic pattern.

TENSES, ASPECTS, AND MOODS

Table 5.19 shows the basic forms of an Arabic verb in the perfective and imperfective aspects. The specific vowel used after consonant ⓪ varies from word to word, but is constant in a particular aspect (column). For example, the second person singular feminine perfective of the root [k-t-b] ‘write’ is [katabti] ‘you have written’, using [a] throughout the perfective. Its corresponding imperfective vowel is [u]. In pausa, many of the final short vowels are omitted. The two aspects are similar to the respective aspects in Russian (§3.5.4.5). In general, the perfective is used for the past. The imperfective is most commonly used for the present, but it can be used for any tense, indicating actions that are viewed as ongoing, not having a definite end. The auxiliary word [sauf], which is usually contracted to [sa-], can be used to make the future meaning of an imperfective explicit.

Table 5.19 shows only the indicative mood, which is the only mood the perfective has. The imperfective form, in contrast, can be modified to express a variety of other moods. The subjunctive, which is used mostly in subordinate clauses expressing purpose, is formed from the imperfective indicative by dropping the

TABLE 5.19

Arabic active indicative verb inflections

Agreement	PFV	IPFV
1SG	①a②V③tu	ʔa①②V③u
2SG.M	①a②V③ta	ta①②V③u
2SG.F	①a②V③ti	ta①②V③i:na
3SG.M	①a②V③a	ja①②V③u
3SG.F	①a②V③at	ta①②V③u
2DU	①a②V③tuma:	ta①②V③a:ni
3DU.M	①a②V③a:	ja①②V③a:ni
3DU.F	①a②V③ata:	ta①②V③a:ni
1PL	①a②V③na:	na①②V③u
2PL.M	①a②V③tum	ta①②V③u:na
2PL.F	①a②V③tun:a	ta①②V③na
3PL.M	①a②V③u:	ja①②V③u:na
3PL.F	①a②V③na	ja①②V③na

Note: V is a short vowel.

final syllable from any two-syllable suffix (i.e. the part added after consonant ③) and changing any final [u] to [a]. Thus the first person singular [ʔaktubu] becomes [ʔaktuba], the third person plural masculine becomes [jaktubu:], and the corresponding feminine is an unchanged [jaktubna]. The JUSSIVE MOOD (JUS), which usually expresses a command or prohibition, can be formed from the subjunctive by further deleting any suffix that consists entirely of [a]; for example, [la: taktub] ‘do not write!’. The imperative is formed from second-person jussive forms by replacing the part before consonant ① with [i]; but if the verb uses [u] after consonant ② throughout the imperfective, that vowel is used for the prefix; for example, [uktub] ‘write!’.

The derived forms in Table 5.19 vary if there are extra consonants, long vowels, and other complications. A particularly striking variant comprises verbs that have a glide as the middle consonant: [w] (24a–b) or [j] (24c–d). In the basic form as well as a few derivations like [a①②a③a], the glide and its surrounding vowels contract to a single vowel. In the perfective (24a,c), that tends to be an [a:] if that vowel would be in an open syllable, but in a closed syllable, the contraction results in a short vowel that corresponds to the glide: [i] from [j], [u] from [w]. In the imperfective (24b,d), the glide disappears without a trace, but the following vowel lengthens in open syllables.

- (24) a. ‘say’ PFV: *[qawala] > [qa:la], *[qawalna] > [qulna]
 b. ‘say’ IPFV: *[jaqwulu] > [jaqu:lu], *[jaqwul] > [jaqul]
 c. ‘travel’ PFV: *[sajara] > [sa:ra], *[sajarna] > [sirna]
 d. ‘travel’ IPFV: *[jasjiru] > [jasi:ru], *[jasjir] > [jasir]

PASSIVE

The passive paradigm of the verb is very similar to the active one except that the first two vowels of the transfix are changed. The perfective starts with [①u②i③-], and the imperfective has [-u①②a③-].

- (25) a. [kataba] PFV.ACT ‘wrote’
 b. [kutiba] PFV.PASS ‘was written’
 c. [jaktubu] IPFV.ACT ‘is writing’
 d. [juktabu] IPFV.PASS ‘is being written’

Passive constructions express the patient role – what would normally be the object – as a subject. This leaves the agent – what would normally be the subject – without a grammatical relation to express it. In English and several other languages, the agent can be rescued by expressing it as the object of a preposition: ‘Zayd was killed by Muḥammad’. In Arabic passive constructions, there is no way for the agent to be expressed. This is not a rare phenomenon among the languages of the world.

PARTICIPLES

The active participle transfix is [①a:②i③-], for example, [ka:tibun] ‘writing’. The passive participle transfix is [ma①②u:③-], for example, [maktu:bun] ‘written’.

5.6.5 SYNTAX

5.6.5.1 Word order

Modern Standard Arabic has a variety of orders, which are used depending on the type of constituents found in the sentence, but generally it is classified as a VSO language (26a). If there is no verb in the sentence – like the verbless sentences in Russian, §3.5.5 – the subject normally comes first (26b). Subject pronouns are optional. If one is used, it usually precedes the verb ([huwa] in 26c).

- (26) a. [za:ra zaidun ʃamran] ‘Zayd visited Amr.’
 b. [zaidun mari:dʕun] ‘Zayd (is) ill.’
 c. [huwa za:ra ʃamran] ‘He visited Amr.’

Adjectives follow the nouns they modify (27a), but determiners precede their heads (27b).

- (27) a. al=luβ-at-u l=ʃarab-ij:-at-u
 DEF=language-F-NOM DEF=Arabic-ADJ-F-NOM
 ‘the Arabic language’

- b. ha:ða: l=luḡ-at-u
 this DEF=language-F-NOM

5.6.5.2 Agreement and predication

Adjectives must agree with the nouns they modify in number, gender, case, and definiteness. Recall that some nouns are feminine even though they are not explicitly marked with [-at]; modifying adjectives must agree with their implicit gender.

- (28) a. busta:n-u-n kabi:r-u-n
 garden-NOM-INDF big-NOM-INDF
 ‘a big garden’
- b. al=busta:n-u l=kabi:r-u
 DEF=garden-NOM DEF=big-NOM
 ‘the big garden’
- c. ad=da:r-u l=kabi:r-at-u
 DEF=house(F)-NOM DEF=big-F-NOM
 ‘the big house’

However, predicative adjectives are indefinite, regardless of the definiteness of the subject:

- (29) ad=da:r-u kabi:r-at-u-n
 DEF=house(F)-NOM big-F-NOM-INDF
 ‘The house is big.’

Adjectives that modify plural (not dual) nouns often take the feminine singular form, especially if the noun does not refer to humans.

- (30) al=mudunu kabi:r-at-u-n
 DEF=cities big-F-NOM-INDF
 ‘The cities are big.’

Possessed nouns are grammatically definite, even though they are not preceded by the definite article, and therefore adjectives modifying them must also be definite:

- (31) kalb-u=hu l=kabi:r-u
 dog-NOM=3SG.M DEF=big-NOM
 ‘his big dog’

5.6.5.3 Relative clauses

If the modified noun is definite, the relative clause is introduced by the relative pronoun [laði:], which is always preceded by the definite marker [al=]. The relative pronoun agrees in gender and number with the antecedent. It also agrees in case, but that is apparent only in the dual.

- (32) ar=raḍzul-u l=laði: qatala ʔab=i:
 DEF=man-NOM DEF=REL kill\PFV.3SG.M father=1SG
 ‘the man who killed my father’

If the relative pronoun [al=laði:] has some function other than subject of the verb in the relative clause, a RESUMPTIVE PRONOUN is required. This generally is expressed as an enclitic pronoun, such as [=hu] in this example. A resumptive pronoun refers back to the modified noun and the relative pronoun.

- (33) ar=raḍzul-u l=laði: qatala=hu ʔab=i:
 DEF=man-NOM DEF=REL kill\PFV.3SG.M=3SG.M father=1SG
 ‘the man whom my father killed’
 Literally, ‘the man that my father killed him.’

When the noun that the relative clause modifies is indefinite, the relative clause omits the relative pronoun:

- (34) raʔai-tu raḍzul-a-n ḍza:ʔa
 see\PFV-1SG man-ACC-INDF come\PFV.3SG.M
 ‘I saw a man who came.’

5.6.6 SAMPLE TEXT

The following text is a short story from *One thousand and one nights*, which is a source of wonderful tales both short and long, familiar to people the world over, especially in their expurgated versions for children. This text and its free English translation are based on Tritton (1943: 23–24).

- (35) كَانَ حَمَارٌ يُسَافِرُ بِخَمْرٍ لَهُ وَمَعَهُ قَرْدٌ.
 ka:na ʧam:a:r-u-n jusar:firu bi=ʧamr-i-n
 be\PFV.3SG.M wine_merchant-NOM-INDF travel\IPFV.3SG.M with=wine-GEN-INDF
 la=hu wa=maʃa=hu qird-u-n
 to=3SG.M and=with=3SG.M monkey-NOM-INDF

‘A wine merchant used to travel with his wine, and there was a monkey with him.’
 [ka:na]. The root is [k-w-n]. The perfective is here used as an auxiliary verb to mark past tense. The aspect is supplied by the imperfective verb that follows the subject [ʧam:a:r-u-n]. If [ʧam:a:r-] were the predicate of [ka:na] (‘he was a wine merchant’), it would be in the accusative case.

[χam:a:r-]. The [a@a:a@a] pattern marks habitual involvement with whatever is expressed by the root. Thus, literally, 'one habitually involved with wine.'

[bi=]. The prepositions [bi=] 'with' and [li=] 'to' before a noun are always written attached to it and are treated as proclitics. Similarly, the coordinating conjunctions [wa=] and [fa=] are always attached to the following word.

[la=hu]. Before an enclitic pronoun, the prepositions are not proclitics but full words, with [a] instead of [i]. This preposition is often used to indicate a possessor.

- (36) وَكَانَ يُمِزُّجُ الْخَمْرَ بِالْمَاءِ نِصْفَيْنِ
 wa=ka:na jamzudʕu l=χamr-a bi=l=mar:?-i nisʕf-aini
 and=be\PFV.3SG.M mix\IPFV.3SG.M DEF=wine-ACC with=DEF=water-GEN half-DU.ACC
 'He would mix the wine with water half and half'

- (37) وَيَبِيعُهُ بِسَعْرِ الْخَمْرِ.
 wa=jabi:ʕu=hu bi=siʕr-i l=χamr-i
 and=\IPFV.3SG.M=3SG.M with=price-GEN DEF=wine-GEN
 'and he would sell it at the price of wine.'
 [jabi:ʕu]. Root is [b-j-ʕ].

- (38) وَأَلْقَرْدُ يُشِيرُ إِلَيْهِ أَنْ لَا تَفْعَلَ فَيَصْرُبُهُ.
 wa=l=qird-u juʕi:ru ?ilai=hi ?an la: tafʕal
 and=DEF=monkey-NOM sign\IPFV.3SG.M to=3SG.M that(SUBORD) NEG do\JUS.2SG.M
 fa=jadʕribu=hu
 so=beat\IPFV.3SG.M=3SG.M
 'The monkey would signal to him: "Don't do that!", and so he would beat him.'
 [juʕi:ru]. Perfective is [aʕa:ra], from root [ʕ-w-r].

- (39) فَلَمَّا قَرَعَ مِنْ بَيْعِ الْخَمْرِ
 fa=lam:a: faraʕa min baiʕ-i l=χamr-i
 and=when finish\PFV.3SG.M from sell-GEN DEF=wine-GEN
 'When he had finished selling the wine,'
 [baiʕi] is a verbal noun formed from [b-j-ʕ].

- (40) وَأَرَادَ الرُّجُوعَ إِلَى بَلَدِهِ رَكِبَ الْبَحْرَ
 wa=?ara:da r=rudʕu:ʕ-a ?ila: balad-i=hi
 and=want\PFV.3SG.M DEF=return-ACC to hometown-GEN=3SG.M
 rakiba l=baʕr-a
 travel\PFV.3SG.M DEF=sea-ACC
 'and wanted to go back to his town, he traveled the sea'
 [?ara:da]. [a@a@a@a] derivative of [r-w-d].
 [rudʕu:ʕ] is a verbal noun.

- (41) وَقَرْدُهُ مَعَهُ وَخُرْجٌ فِيهِ ثِيَابُهُ
 wa=qird-u=hu maʕa=hu wa=χurd̪ʒ-u-n fi:=hi
 and=monkey-NOM=3SG.M with=3SG.M and=saddle_bag-NOM-INDF in=3SG.M
 θija:b-u=hu
 clothes-NOM=3SG.M
 ‘and his monkey was with him, and a saddle bag in which were his clothes’
 [θija:b-] is the plural of [θaub].
- (42) وَأَلْكَيْسُ الَّذِي جَمَعَهُ مِنْ مِّمَّنِ الْخَمْرِ.
 wa=l=ki:s-u l=laði: d̪ʒamaʕa=hu min θaman-i l=χamr-i
 and=DEF=purse-NOM DEF=REL collect\PFV,3SG.M=3SG.M as price-GEN DEF=wine-GEN
 ‘and the purse that he had collected for the price of the wine.’
- (43) فَلَمَّا سَارَ فِي الْبَحْرِ اسْتَخْرَجَ الْقِرْدُ الْكَيْسَ مِنْ مَوْضِعِهِ
 fa=lam:a: sa:ra fi: l=baħr-i staχrad̪ʒa
 and=when be_underway\PFV,3SG.M in DEF=sea-GEN pull_out\PFV,3SG.M
 l=qird-u l=ki:s-a min mawdʿif-i=hi
 DEF=monkey-NOM DEF=purse-ACC from place-GEN=3SG.M
 ‘When he was underway at sea, the monkey pulled the purse from its place’
 [sa:ra]. From root [s-w-r].
 [staχrad̪ʒa]. Root [χ-r-d̪ʒ] ‘go out’.
- (44) وَرَقِيَ الدَّقْلَ وَهُوَ مَعَهُ حَتَّى صَارَ فِي أَعْلَاهُ.
 wa=raqija d=daqal-a wa=huwa maʕa=hu ha:ta:
 and=climb\PFV,3SG.M DEF=mast-ACC and=3SG.M with=3SG.M until
 sʕa:ra fi: ʔaʕla:=hu
 arrive\PFV,3SG.M in top=3SG.M
 ‘and climbed the mast, it being with him, till he reached its top’
 [sʕara]. Root is [sʕ-j-r].
- (45) وَرَمَى إِلَى الْمَرْكَبِ بِدِرْهَمٍ
 wa=rama: ʔila: l=markab-i bi=dirham-i-n
 and=throw\PFV,3SG.M into DEF=ship-GEN with=dirham-GEN-INDF
 ‘He threw a dirham into the ship’
 [rama:]. The root is [r-m-j]. In most forms of such verbs the final glide shows up as such or as the corresponding vowel [i], but this is one of the forms in which there is a contraction to [a:], much as was seen in verbs that have a glide as the second root consonant. This verb obligatorily requires a prepositional phrase headed by [bi] to show the object that was thrown.
 [markab-]. This is a locative noun form of the verb root [r-k-b] ‘to travel’.

(46)

وَإِلَى الْبَحْرِ بِدِرْهَمٍ

wa=?ila: l=bahr-i bi=dirham-i-n
 and=into DEF=sea-GEN with=dirham-GEN-INDF
 ‘and a dirham into the sea’

(47)

فَلَمْ يَزَلْ ذَلِكَ دَابُّهُ حَتَّى قَسَمَ الدَّرَاهِمَ نِصْفَيْنِ.

fa=lam jazal ḏa:lika da?ba=hu hat:a: qasama
 and=NEG cease\JUS.3SG.M that(M.SG) persistence=3SG.M until divide\PFV.3SG.M
 d=dara:him-a nis^f:f-aini
 DEF=dirham\PL-ACC half-DU.ACC

‘and he didn’t stop what he was doing until he had divided the dirhams half and half.
 [jazal] has the root [z-j-l].

5.7 Sketch of Swahili

5.7.1 GENETIC AFFILIATION AND GENERAL BACKGROUND

Swahili belongs to the Bantu subbranch of the East Benue-Congo branch of the Niger-Congo language family. Reports of the number of speakers of Swahili differ widely, in part because a great majority of the people who use it are native speakers of other languages; their fluency in Swahili varies widely.

Originally a language spoken by a relatively small number of people living on the east coast of Central Africa, Swahili became a trade language and a lingua franca, probably because some kind of intermediary language was needed between the Arab traders and the local indigenous population. The name of the language, *Swahili*, is derived from the Arabic word for ‘coastal’. From the coast it spread inland and seaward, and now it is an official or widely spoken language in Tanzania, Kenya, Uganda, the Comoro Islands, and the Democratic Republic of the Congo, and is also used in parts of Mozambique, Somalia, and Madagascar.

Four major groups of dialects are found on the coast:

- ❑ Kiunguja was originally spoken on the island of Zanzibar and is now also spoken on the African mainland in Tanganyika. Both regions constitute the state of Tanzania. Modern standard Swahili, the subject of this sketch, is closest to this dialect.
- ❑ Kimvita is spoken on Mombasa Island and other parts of Kenya.
- ❑ Kiamu is spoken on the island of Lamu (northern coast of Kenya) and the coastal areas opposite that island.
- ❑ Kimwani is spoken in northern Mozambique and the Kerimba Islands.

Originally, Swahili was written in Arabic script, but starting in the nineteenth century the Latin script replaced it.

5.7.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

The consonants are listed in Table 5.20. Where the Standard Swahili orthographic equivalents differ from the IPA, they are cited in angled brackets.

The interdental and velar fricatives are found only in Arabic loanwords. Many Swahili speakers substitute other sounds for them. For example, [x] ⟨kh⟩ is often pronounced [h], and [ɣ] ⟨gh⟩ as [g].

As in Tibetan transcriptions (§4.11.2), the symbol [ʰ] stands for prenasalization on any consonant. The voiced oral stops that are not prenasalized are implosives, sounds that we encountered in other African languages, including Hausa (§5.1.3). In addition to having prenasalized stops like [ʰb], Swahili also has sequences of nasal consonants followed by implosive phonemes, as in [mʙ]. Unfortunately, in ordinary orthography, the two are spelled the same, ⟨mb⟩. For example:

- (48) a. ⟨mtu mbaya⟩ [m^{tʰ}u mʙaja] ‘a bad person’
 b. ⟨nyumba mbaya⟩ [ɲuⁿba ⁿbaja] ‘a bad house’

Swahili has syllabic nasals. In general, nasal phonemes are syllabic before any consonant other than a glide or voiced fricative. Thus, in (48a), both [m]s are syllabic (49a–b). [m^{tʰ}u] is a two-syllable word and [mʙaja] is a three-syllable word. Another example: [ɲt̩i] ‘country’ (49c). Perhaps more surprisingly, prenasalized stops can also be syllabic. Like many languages, Swahili has a MINIMAL-WORD CONSTRAINT, meaning that lexical words, such as nouns, cannot be shorter than a certain length. In Swahili, lexical words have to have at least two syllables. If a word would otherwise have one syllable, a prenasalized stop becomes its own syllable (49d). Note that unlike a phonemic sequence ^x[mʙwa], a syllabic prenasalized stop is pronounced with a plosive, not an implosive, stop. Although we will call attention in this sketch to some important instances where syllabic nasals

TABLE 5.20

Consonants of Swahili

p	t	ʃ̣ ⟨ch⟩	k	
p ^h ⟨p⟩	t ^h ⟨t⟩	ʃ̣ ^h ⟨ch⟩	k ^h ⟨k⟩	
ʙ ⟨b⟩	ɗ ⟨d⟩	f ⟨j⟩	ɣ ⟨g⟩	
ⁿ b ⟨mb⟩	ⁿ d ⟨nd⟩	ⁿ ɗ̣̣̣ ⟨nj⟩	ⁿ ɣ ⟨ng⟩	
f	θ ⟨th⟩	s	ʃ ⟨sh⟩	x ⟨kh⟩ h
v	ð ⟨dh⟩	z		ɣ ⟨gh⟩
m	n	ɲ ⟨ny⟩	ŋ ⟨ng⟩	
	r			
	l			
		j ⟨y⟩	w	

are common, we will treat this phenomenon as allophonic and will not normally notate syllabicity in our phonemic transcriptions.

- (49) a. /mtu/ [mtu]
 b. /mbaja/ [mbaja]
 c. /ɲtʃi/ [ɲtʃi]
 d. /ⁿbwa/ [mbwa]

Aspirated stops are not terribly common in Swahili, because they arise only when a nasal used to precede an unaspirated stop (50). Because the phonetic condition – the nasal – has disappeared, the contrast between aspirates and unaspirated stops is now phonemic: [paa], for example, means ‘roof’. The orthography, however, spells both words the same way, with a ⟨p⟩.

- (50) [*npaa] > [*ɲpaa] > [*hpaa] > [p^haa] ‘gazelle’

Swahili has a typologically common triangular five-vowel system, contrasting [i], [u], [e], [o], and [a]. It has traces of an assimilation rule based on vowel height. Certain suffixes that contain the vowel [i] lower it to [e] after stems that contain mid vowels (51a–b), but not after high (51c) or low (51d) vowels:

- (51) a. [end-] ‘go’, [end-e-] ‘go to’
 b. [som-] ‘read’, [som-e-] ‘read for’
 c. [pig-] ‘hit’, [pig-i-] ‘hit for’
 d. [faj-] ‘do’, [faj-i-] ‘do for’

Stress is on the PENULTIMATE – next-to-the-last – syllable.

5.7.3 MORPHOLOGY

Typologically, Swahili is very agglutinative, although it also has some fusional traits, such as irregular or morphologically conditioned allomorphs of some morphemes. It has a number of suffixes, but, like most Bantu languages, it is predominantly a prefixing language, a situation that is much less common crosslinguistically than a predominance of suffixes.

5.7.3.1 Nouns and their modifiers

The major inflection on nouns consists of prefixes that indicate the noun’s gender and number; the two categories are expressed in a fusional fashion by a single morpheme. Adjectives and determiners as well as verbs governing nouns have to show gender and number agreement with the noun. For example:

- (52) wa-t^hu wa-zuri wa-wili wa-le wa-meaⁿguka
 AN.PL-person AN.PL-good AN.PL-two AN.PL-that AN.PL-fell_down
 ‘Those two good people fell down.’

In the gloss, AN stands for the animate gender. Because the subject of the sentence in the preceding example is an animate noun, the verb, too, must have the [wa-] prefix indicating agreement with the class of the subject noun. Agreement markers on determiners and verbs are often identical to the ones on nouns, but in some genders they may differ.

Modifiers and determiners follow their noun, and there is a preferred order of such elements, as illustrated in (52). For a more detailed statement on word order, see the section on syntax (§5.7.4).

Gender in Swahili is often called *class*. There are some prominent differences between Swahili gender and that of the Indo-European and Afro-Asiatic languages we have looked at. Swahili has seven genders, whereas Russian has three (§3.5.4.1) and Arabic only two (§5.6.4.1). Whereas the latter languages have a masculine and feminine gender into which names for men and women are fairly reliably classified, Swahili gender has nothing to do with biological sex. Nevertheless the more abstract principles are the same: nouns are divided into classes, and certain words connected with them must have markers that denote the same class as the noun. Table 5.21 lists the genders of Swahili. It gives a name to each gender and tells what pronominal forms are used in reference to nouns of that gender. The table also gives some indication of the characteristic meaning of nouns that have each gender. But, as is typical of gender systems in other languages, most Swahili genders are not semantically homogeneous. Like the masculine and feminine genders of Russian, which contain innumerable nouns naming objects that are neither male nor female, most genders in Swahili contain many words whose membership defies explanation.

Most nouns in Swahili take prefixes that indicate their own gender and number (Table 5.22). Included in the table are the numbers assigned by Meinhof (1906) in his comparative study of the Bantu languages. Note that Meinhof gave separate numbers to the singular and plural forms of each gender.

TABLE 5.21
Swahili genders

Gender		Pronoun	
		SG	PL
ABST	abstract	u	ja or Ø
AGGR	aggregate	u	zi
AN	animate	a	wa
AUG	augmentative	li	ja
DIM	diminutive	ki	vi
N	neuter	i	zi
VIT	vital	u	i

TABLE 5.22

Swahili gender markers on nouns

Gender	SG	PL
ABST	NC14 u-	NC6 ma-
AGGR	NC11 u-	NC10 ⁿ -, ∅, ^h -, ⁿ -
AUG	NC5 fi-, f-, ∅	NC6 ma-
AN	NC1 mw-, m-	NC2 wa-, w
DIM	NC7 ki-	NC8 vi-
N	NC9 ⁿ -, ∅, _h	NC10 ⁿ -, ∅, ^h -
VIT	NC3 mw-, m-	NC4 mi-

Note: NC: Meinhof noun class number.

ANIMATE

As we have seen in Russian (§3.5.4.1) and Burushaski (§4.8), animacy is an important category in many languages. In Swahili, the animate gender is most consistently associated with the prefix pair [mw-], plural [wa-]. Almost all nouns that take those prefixes name humans, although the reverse is not necessarily true: names for humans and other animate beings also appear with other prefixes. [mw-] loses its [w] before [u] and consonants, usually ending up as a syllabic [m̩]. The plural prefix is [wa-], but with a following stem vowel may contract to [wa-] (with [a-] stems) or [we-] (with [i-] or [e-]). In general, the same prefixes that are used for nouns are also used for adjectives, but for various reasons the prefixes in any given noun phrase might not all be identical. For phonetic reasons, a noun might have syllabic [m̩] for a prefix but be modified with an adjective that has [mw-], or vice versa.

- (53) a. [mt^hu] ‘person’, [wat^hu] ‘people’
 b. [mɔduɔ] ‘insect’, [wafuɔ] ‘insects’

VITAL

All nouns of the VITAL GENDER have in the singular the same prefix as for those used for the human animate nouns, but the plural prefix is [mi-]. The vital gender contains words naming trees and other plants (54a), but note that the augmentative gender is used for most names of fruit (55d). In addition, the vital gender contains nouns that refer to various products made of wood, names of some natural phenomena and entities (54b–d), and names of some parts of the body (54e–f).

- (54) a. [mti] ‘tree’, [miti] ‘trees’
 b. [mto] ‘river’, [mito] ‘rivers’
 c. [mlima] ‘mountain’, [milima] ‘mountains’
 d. [mwezi] ‘moon’, [miezi] ‘moons’
 e. [mojo] ‘heart’, [miojo] ‘hearts’
 f. [mɟuu] ‘leg’, [miɟuu] ‘legs’

AUGMENTATIVE

The singular prefix for the AUGMENTATIVE GENDER may have various shapes: [f-] before vowels, [fi-] before consonants, or, most often, no prefix at all. The plural is formed with [ma-], which usually is prefixed to any existing [fi-], rather than replacing it. The name of the gender comes from the fact that it often names large things. Such names may be derived from words of other genders simply by changing the prefix and agreement patterns (55a). The augmentative gender is also used for fruit and other round objects (55b–c). Names of fruit typically pair with names of trees, which are in the vital gender (55d).

- (55) a. [mto] VIT 'river', [fito] AUG 'big river'
 b. [jai] 'egg'
 c. [ziwa] 'lake'
 d. VIT [m̄t̄juⁿgwa] 'orange tree', [m̄it̄juⁿgwa] 'orange trees'; AUG [t̄juⁿgwa] 'orange', [mat̄juⁿgwa] 'oranges'

DIMINUTIVE

The prefix for the DIMINUTIVE GENDER is [ki-] before consonants and [t̄j-] before vowels other than [i]. In the plural, [vi-] is used, but [vj-] before vowels. The name of the gender comes from the fact that it often names small things, and diminutive counterparts to other words can be derived by replacing their prefix, if any, with [ki-] or [ki-fi-]. The diminutive gender is also extensively used for names of tools, small useful things, physical ailments or people with disfigurements, and languages. Both the augmentative and diminutive prefixes can be applied to words for humans and animals; such words retain animate agreement, unless disparagement is intended.

- (56) a. [mto] VIT 'river', [kifito] DIM 'stream'
 b. [kitabu] 'book'
 c. [kilema] 'a lame person'
 d. [kiswahili] 'Swahili language'

The Arabic loanword [kitabu] was reanalyzed by its Swahili borrowers as consisting of the diminutive prefix [ki-] followed by a root [tabu], even though in Arabic the [k] is part of the root, [k-t-b] (§5.6.4.1).

NEUTER

The prefix for the neuter gender is quite varied. It can be thought of as basically [ɲ], the form it takes before vowels (57a). Before consonants, the nasal assimilates to the place of articulation. With voiced stops, it makes a prenasalized stop (57b). To fulfil the minimal-word requirement, if the resultant word is only one syllable, the prenasalized stop is syllabic (57c). Before consonants in other monosyllabic

roots, the nasal becomes a syllabic nasal with the place of articulation of the next consonant (57d). Before voiceless stops in polysyllabic roots, the nasal combines with that consonant to make an aspirated stop (50, 57e); note that each vowel in Swahili is always in its own syllable, so that [-paa] counts as a disyllabic root. In other environments, the nasal manifests as the nasal HOMORGANIC to the next consonant – that is, having the same place of articulation. The plural prefix is the same as the singular on nouns and adjectives.

- (57) a. [-oka]: [ɲoka] ‘snake’
 b. [-buzi]: [ʰbuzi] ‘goat’
 c. [-ɓwa]: [ʰbwa] ‘dog’
 d. [-pwa]: [mpwa] ‘nephew’ or ‘niece’
 e. [-paa]: [pʰaa] ‘gazelle’

Most animal names and family names take the neuter prefix, but take animate gender agreement. The neuter gender proper contains some fruit names, most loanwords – most of which take no prefix – and other miscellaneous groups of nouns, such as names of musical instruments, parts of the body, buildings and their parts, and some countries and other parts of the world.

AGGREGATE

The singular prefix for the AGGREGATE GENDER is [u-], usually [w-] before a vowel. Except for that prefix, aggregate nouns behave just like vital singular nouns when it comes to agreement markers; even adjectives that agree with aggregate nouns have the prefix [m-]. The plural, both in form and agreement, is the same as in the neuter plural, sometimes taking no prefix at all.

Typical nouns having this gender denote mass nouns:

- (58) a. [udɔʰgo] ‘soil’
 b. [uɟali] ‘porridge’
 c. [uʰga] ‘flour’

Some nouns belonging to this gender are names of objects that refer to single units from a group or collection of such objects:

- (59) [uʃaʰga] ‘bead’ (cf. [ʃaʰga] ‘beads’)

The aggregate gender also contains nouns referring to long or elongated objects and some natural phenomena:

- (60) a. [uzi] ‘thread’
 b. [upepo] ‘wind’

ABSTRACT

Nouns of the ABSTRACT GENDER are like those of the aggregate gender in the singular. In the plural, they are like augmentative nouns. Many words of this gender are abstract nouns, often derived from other nouns or from adjectives (61a–b). Many names of countries are also of the abstract gender (61c–d).

- (61) a. [zuri] ‘beautiful’, [uzuri] ‘beauty’
 b. [mofa] ‘one’, [umofa] ‘oneness, unity’
 c. [uɟaⁿda] ‘Uganda’ (literally ‘Ganda country’)
 d. [uiⁿgereza] ‘England’; compare the same root with other gender prefixes:
 [kiiⁿgereza] DIM ‘English language’, [waiⁿgereza] AN ‘English people’

The prefix [ma-] also appears on words for mass nouns like [mafi] ‘water’, and on some nouns derived from verbs, such as [mapeⁿdo] ‘love’. These tend to be words for which it is unusual to distinguish a singular from a plural.

LOCATIVE

Nouns may take the locative suffix [-ni] to express a general locative function, such as [ʃaⁿbani] ‘in the field’. More specific locational information can be added by changing the prefix on a determiner such as a possessive. These determiners distinguish definite locations from indefinite locations, the latter including goals of motion. Here are a couple of examples from Myachina (1981: 32):

- (62) a. kuna wa-tu ʃaⁿba-ni p-aⁿgu
 there_are AN.PL-person field(AUG)-LOC LOC.DEF-1SG
 ‘There are people in my field.’
 b. ni-na-eⁿda ʃaⁿba-ni kw-aⁿgu
 1SG-PRS-go field(AUG)-LOC LOC.INDF-1SG
 ‘I am going to my field.’

5.7.3.2 Possessive constructions

To express possession, Swahili places the head noun (possessed item) first, followed by a possessive marker, then the specifier noun (possessor). The possessive marker takes a prefix that agrees in number and gender with the head noun.

- (63) ki-taⁿbu t̃j̃-a m-toto
 DIM-book DIM-POSS AN-child
 ‘the child’s book’

Possessive pronouns – ‘my’, ‘your’, and so on – are combinations of the possessive marker and a special person marker:

- (64) wa-toto w-a-n-gu
 AN.PL-child AN.PL-POSS-1SG
 ‘my children’

5.7.3.3 Verb morphology

The verb morphology of Swahili is very complex: long strings of prefixes may precede a verb root, and a moderate number of suffixes, some of which are derivational, may follow it. Only a very cursory outline will be given here. To simplify the presentation, all references to allomorphs will be omitted.

The least inflected form of the Swahili verb, as in most languages, is the imperative singular form:

- (65) a. [soma] SG ‘read!’
 b. [someni] PL ‘read!’

Prefixes are attached to the stem in this order:

- (66) subject + tense/aspect/mood + relative + object

SUBJECT PREFIX

This is a pronominal form showing agreement with the gender and number of the subject of the verb. The third-person forms are listed in Table 5.21. The subject forms for the first and second person are:

- (67) a. [ni-] 1SG
 b. [tu-] 1PL
 c. [u-] 2SG
 d. [m(w)-] 2PL

TENSE/ASPECT/MOOD PREFIX

The next prefix expresses a combination of tense, aspect, and mood.

- (68) a. [-a-] indefinite, general time
 b. [-na-] present
 c. [-li-] past
 d. [-me-] perfective
 e. [-ta-] future

- f. [-ka-] SEQUENTIAL MOOD (SEQ) ‘and then’ (an event that follows the one just portrayed)
- g. [-ki-] conditional (“if you see them, say hello”)
- h. [-ⁿge-] hypothetical conditional (“if this fell, it would break”)
- i. [-ⁿga-li-] conditional CONTRAFACTUAL MOOD (contrary to what actually happened) (“if this had fallen, it would be broken”)

- (69) a. ni-na-soma
1SG-PRS-read
‘I am reading.’
- b. ki-su ki-mofa ki-ta-tofa
DIM-knife DIM-one DIM-FUT-suffice
‘One knife will be enough.’

The habitual prefix [hu-] is always used without a subject prefix. The tense of habitual verbs is to be inferred from the context.

Negative prefixes entail special complications, which will be discussed separately.

RELATIVE PREFIX

If the verb is in a relative clause, the next prefix is a pronominal agreeing in number and gender with the head noun governing the relative clause. This pronoun is followed by an [o]. In (70a), the [t̃] is the form [ki] takes before a vowel; this is the pronoun that agrees in number and gender with the head noun [kitabu]. Exceptionally, the animate singular relative prefix is [je-] (70b).

- (70) a. ki-tabu a-li-t̃-o-ki-soma m-toto
DIM-book AN-PST-DIM-REL-DIM-read AN-child
‘the book that the child read’
- b. m-toto a-li-je-soma ki-tabu
AN-child AN-PST-AN.REL-read DIM-book
‘the child who read a book’

There are various complications with such relative constructions, in that the order of the morphemes is different in different tenses, and there is a special negative construction in which tense marking is neutralized.

There are special relatives of place as well:

- (71) ha-tu-fui a-li-k-o-kwenda
NEG-1SG-know AN-PST-LOC.INDF-REL-go
‘We don’t know where he went.’

OBJECT PREFIX

Next comes a pronoun agreeing in number and gender with the object of the verb. The pronoun agreement is optional for inanimate objects; if used, it generally indicates that the direct object is definite or bears special emphasis:

- (72) a. ni-li-soma ki-tabu
 1SG-PST-read DIM-book
 'I read a book.'
- b. ni-li-ki-soma ki-tabu
 1SG-PST-DIM-read DIM-book
 'I read the book.'

Usually the object prefixes are selected from the same set as the subject prefixes (§5.7.3.3), but there are special forms for agreeing with the following animate categories:

- (73) a. 2SG [ku-]
 b. 2PL [wa-]
 c. 3SG [mw-] ([m-] before consonants)

Swahili verbs also take various suffixes, in the following order. Of these, the only one that must appear on all verbs is the final one, which expresses mood:

- (74) causative + applicative + passive + mood

CAUSATIVE SUFFIX

The causative is marked by an underlying suffix *[-j], which surfaces as palatalization of the preceding consonant. Thus, [k] becomes [ʃ]; [g], [d], and sometimes [b] become [z]; and so on. However, most verbs use the fuller suffix [-iʃ], which may become [-eʃ] by assimilation (§5.7.2).

APPLICATIVE SUFFIX

The APPLICATIVE VOICE (APPL) is usually marked by the suffix [-i], or [-e] by assimilation:

- (75) a-li-ni-aⁿdik-i-a barua
 AN-PST-1SG-write-APPL-IND letter(N)
 'He wrote me a letter.'

As can be seen from this example, this prefix indicates that a role that would not ordinarily be subject or object of the verb has been made the direct object. Thus the 1SG prefix [ni-] refers not to the thing that was written, which is neuter, but to the person to whom the letter was written.

PASSIVE SUFFIX

The passive voice is marked by the suffix [-w].

MOOD SUFFIX

Verbs are usually encountered with the ending [-a], which can be thought of as a marker of the indicative mood non-negative. Negatives have other endings (§5.7.3.3), and the subjunctive mood ends in [-e].

Besides the above affixation, there are many derivational patterns, some involving other suffixes and others involving such things as REDUPLICATION (76d). Reduplication consists of repeating all or part of a base word; the linguistic notation for this is “~”:

- (76) a. [fuⁿb-a] ‘close’
 b. [fuⁿb-u-a] ‘open’: opposite
 c. [fuⁿb-ik-a] ‘be closed’: STATIVE ASPECT (STAT): showing a steady state instead of an action
 d. [fuⁿb-a~fuⁿb-a] ‘keep on closing’: ITERATIVE ASPECT (repetition)
 e. [fuⁿb-an-a] ‘close together’: RECIPROCAL VOICE (RECP) (both participants play the same role with respect to each other)

‘BE’ AND ‘HAVE’

The verb ‘be’ is irregular in Swahili. Its root is [wa] (77a), which, however, does not appear in all the forms of this verb. Instead, another root [li] is used (77b), and even that is totally dropped in particular environments (77c). [ku], which otherwise appears on infinitives, is usually added to verb roots that consist of a single syllable (77a), though not when a relative marker appears at the end of the verb (77b).

- (77) a. u-li-ku-wa
 2SG-PST-INF-be
 ‘you were’
 b. u-li-je
 2SG-be-AN.REL
 ‘you who are’
 c. ni
 be(1SG)
 ‘I am’

There is also an emphatic (EMP) copula construction:

- (78) ⁿdi-mi
 COP(EMP)-1SG
 ‘It is I.’

The suffix [-mi] is a form of the independent pronoun [mimi] ‘I’.

There are also forms like [ni-p-o] ‘I am here’, which consist of a personal prefix followed by [po], which is the locative pronoun [pa] followed by the relative marker [o]. The form [ni] or its negative counterpart [si] can be used without subject agreement when ‘be’ is used simply as a copula.

‘Have’ is expressed by a combination of ‘be’ followed by the comitative marker [na]:

- (79) a. u-li-ku-wa na feḏa
 2SG-PST-INF-be COM money(N)
 ‘Did you have money?’
 Literally, ‘Were you with money?’
- b. ki-ḃaⁿda ki-na vj-uⁿba vi-tatu
 DIM-hut DIM-COM DIM.PL-room DIM.PL-three
 ‘The hut has three rooms.’

In (79b), the verb ‘be’ has a zero allomorph, and therefore the subject agreement prefix is attached to [na] directly.

To express existence, locative prefixes are added to the verb ‘be’ (which is zero in this example), followed by [na]:

- (80) ku-na wa-t^hu ⁿḏḓia-ni
 LOC.INDF-COM AN.PL-person road(N)-LOC
 ‘Are there people on the road?’

NEGATION

Negation is signaled by the prefixes [ha-] and [si-]. The latter appears in place of [ha-ni] (NEG-1SG) and in such constructions as the relative, subjunctive, and conditional:

- (81) a. h-a-ta-ku-wa
 NEG-AN-FUT-INF-be
 ‘He won’t be.’
 The [a] of [ha-] is deleted before another [a].
- b. si-w-i
 NEG(1SG)-be-NEG
 ‘I am not.’

Note that in (81b) the final vowel of the verb stem [wa] is changed from [a] to [i]. This happens only in the negative forms of the present and general tenses.

Tense is neutralized in negative relative constructions:

- (82) t̥j-akula ki-si-t̥j-o-toʃe
 DIM-food DIM-NEG-DIM-REL-sufficient
 ‘food which is/was/will be insufficient’

Both [li-] past and [me-] perfective are replaced by [ku-] in negative forms of the verb.

- (83) h-a-ku-soma
 NEG-3SG-PST-read
 ‘He did not/has not read.’

INFINITIVE

Infinitives are formed by prefixing [ku-] before consonants and [kw-] before vowels; for example, [kusoma] ‘to read’. They can be modified by adjectives, which show agreement with the infinitive by also taking the [ku-] prefix.

5.7.4 SYNTAX

5.7.4.1 Word order

The basic word order in Swahili is SVO, and modifiers generally follow what they modify. Declarative and interrogative sentences have the same word order. Interrogative sentences differ from declarative ones only in intonation or by containing question words.

The order of various nominal modifiers is a mirror image of the usual order of such elements in English:

- (84) noun + adjective + number + determiner

Thus ‘my two big baskets’ is literally ‘baskets big two my’ in Swahili: [vikapu vikuḃwa viwili vja^agu].

5.7.4.2 Grammatical relations

There are two syntactic hints as to which noun is the subject and which is the object of a verb. The first is word order: subjects generally precede objects. The other clue is the agreement morphemes in the verb. There, the subject pronoun must come before the tense marker, and the object pronoun must follow it. Since those pronouns agree with the noun in gender and number, then, as long as the subject and object are of different genders or numbers, the pronouns will indicate which noun is subject and which is object. This allows some freedom in the Swahili word order while avoiding ambiguity:

- (85) ki-taḃu wa-na-ki-som-a wa-toto
 DIM-book AN.PL-PRS-DIM-read-IND AN.PL-child
 ‘The children are reading the book.’

The order of prefixes in the verb clearly shows that the subject of (85) must be whichever noun is animate plural, and the object must be a noun that is diminutive singular.

The only case affix is the locative [-ni], which can be attached to nouns belonging to most genders (§5.7.3.1). Other locative and instrumental relations are signaled by prepositions. For example, the agent noun phrase in a passive sentence is preceded by [na]:

- (86) m-toto a-li-pig-w-a na baḃa j-ake
 AN-child AN-PST-beat-PASS-IND COM father(AN) AN-his
 ‘The child was beaten by his father.’
 [baḃa] has neuter morphology but animate agreement.

As we have already mentioned (§5.7.3.3), a recipient or beneficiary is signaled by an applicative suffix added to the verb, with the verb’s object prefix showing agreement with the noun expressing that relation. Indirect object nouns precede direct object nouns in the verb phrase.

5.7.4.3 Agreement

As already discussed in the morphology section (§5.7.3), noun modifiers and determiners must agree in number and gender with their head noun. In addition, verb forms, too, are inflected to show agreement with their subjects and sometimes also with their objects. If a sentence has several subject nouns belonging to different classes, the subject concord prefix on the verb agrees with the class of the subject noun closest to the verb.

5.7.4.4 Relative and other subordinate clauses

Because Swahili has such a rich inflectional morphology, its syntax is relatively uncomplicated. For example, as was already discussed under verb morphology (§5.7.3), there is a special affix that makes a relative form of the verb; thus, there is no need for such things as relative pronouns to signal relative clauses:

- (87) a. [wa-li-som-a] ‘they read’
 b. [wa-li-o-soma] ‘who read’

However, there is another way to make a relative clause: by using [aⁿba]. This was originally a verb meaning ‘to say’, but it is now a frozen form that acts very much like an English relative pronoun:

- (88) huju ni m-toto aⁿba-je keḃo ha-ta-kufa juḃe
 this is AN-child REL.PRO-REL.AN tomorrow NEG-FUT-come school(N)
 ‘This is the child who will not be coming to school tomorrow.’

Note that [aⁿba] takes the relative suffix that agrees with the number and gender of the modified noun.

Other types of subordinate clauses are introduced by conjunctions. Subordinate clauses that express purpose or intention may or may not be introduced by a subordinate conjunction such as [ili] or [kusufi]. In either case, the verb must be in the subjunctive mood:

- (89) wa-li-oⁿdok-a kusufi wa-eⁿd-e bara hiⁿdi
 AN.PL-PST-leave-IND in_order_to AN.PL-go-SBJV mainland(N) India
 ‘They left in order to go to India.’

5.7.5 SAMPLE TEXT

The following short texts are taken from Perrott (1951: 139), who in turn got them from the Swahili language newspaper *Maendeleo* ‘Progress’. They constitute two short letters from different readers, one arguing against the use of Swahili and the other arguing for it. Such arguments were frequent in East Africa just before independence. The free translations are adapted from Perrott’s.

- (90) KiSwahili Kiondoshwe
 ki-swahili ki-oⁿdoj-w-e
 DIM-Swahili DIM-go_away\CAUS-PASS-SBJV
 ‘Let’s Get Rid of Swahili’
 [oⁿdoj-] is the causative of [oⁿdok-].
- (91) Maneno mengi yatumikayo katika lugha ya KiSwahili yametoka katika lugha za WaArabu, WaZungu na WaHindi.
 ma-neno m-eⁿgi ja-tum-ika-j-o katika luya
 AUG.PL-word AUG.PL-many AUG.PL-use-STAT-ABST.PL-REL in language(N)
 j-a ki-swahili ja-me-tok-a katika luya z-a
 N-of DIM-Swahili AUG.PL-PFV-go_out-IND in language(N)[pl] N.PL-of
 wa-arabu wa-zuⁿgu na wa-hiⁿdi
 AN.PL-Arab AN.PL-European and AN.PL-Indian
 ‘Many words that are used in Swahili have originated in the languages of Arabs, Europeans, and Indians.’
- (92) KiSwahili kilipatikana hasa katika sehemu za pwani, nacho ni lugha ya kibiashara.
 ki-swahili ki-li-pati-ka-na hasa katika sehemu z-a
 DIM-Swahili DIM-PST-obtain-STAT-COM especially in portion(N)[pl] N.PL-of
 pwani na-t̃j-o ni luya j-a ki-biafara
 coast(N) and-DIM-REL be language(N) N-of DIM-commerce
 ‘Swahili is found especially in parts of the coast, and it is a trade language.’
 [nat̃jo]. Literally translated, this sequence means ‘and which.’

- (93) Hakuna watu ambao kabila lao au taifa lao ni la KiSwahili.

ha-kuna wa-t^hu aⁿba-o kabila la-o au taifa
 NEG-exist AN.PL-person REL.PRO-REL(AN.PL) tribe(AUG) AUG-their or nation(AUG)
 la-o ni la-a ki-swahili
 AUG-their be AUG-of DIM-Swahili

‘There are no people whose tribe or nation can be characterized as being “Swahili.”’

[hakuna]. The negative prefix [ha-] does not co-occur with subject prefixes; such constructions are impersonal.

[ni la ki-swahili]. The Swahili construction ‘be’ + possessive marker with concord + Y has the meaning ‘X is characterized or characterizable as Y.’

- (94) Ingefaa tuiondoshe lugha hiyo sasa kabila lugha zetu wenyewe hazijaharibiwa.

i-ⁿge-faa tu-i-oⁿdoſ-e luya hi-jo sasa
 N-COND-be_proper 1PL-N-go_away\CAUS-SBJV language(N) aforementioned-N now
 kabila luya z-etu w-^enewe ha-zi-fa-haribi-wa
 tribe(AUG) language(N)[pl] N.PL-our AN.PL-self NEG-N.PL-yet-destroy-PASS

‘It would be good if we got rid of this language now while our own tribal languages have not yet been destroyed.’

Here is the opposing article:

- (95) Ubaya wa KiSwahili ni nini?

u-^baja w-a ki-swahili ni nini
 ABST-bad ABST-of DIM-Swahili be what

‘What’s so Bad About Swahili?’

- (96) Kitaondoshwa kwa sababu ni mchanganyiko wa maneno ya KiArabu, KiZungu, KiHindi, na KiBantu?

ki-ta-oⁿdoſ-wa kwa sababu ni m-^tſaⁿgaſ-ik-o
 DIM-FUT-go_away\CAUS-PASS for reason(N) be VIT-mix-STAT-NOUN
 w-a ma-neno j-a ki-^arabu ki-zuⁿgu ki-hiⁿdi
 VIT-of AUG.PL-word AUG.PL-of DIM-Arab DIM-Europe DIM-Hindi
 na ki-^bantu
 and DIM-Bantu

‘Shall it be got rid of because it is a mixture of Arabic, European, Hindi, and Bantu words?’

- (97) Hii si sababu hata kidogo.

hii si sababu(N) hata ki-^dogo
 this(N) be.NEG reason even ADV-small

‘That is no reason at all.’

[ki-^dogo]. The [ki-] prefix does not mean that this adjective modifies some diminutive noun. In this case the diminutive prefix signals adverbial usage: ‘at all, even a bit’. It is, however, more usual to derive adverbs from adjectives by adding a [vi-] prefix.

- (98) Hata KiIngereza ni mchanganyiko wa KiLatini na lugha nyingine za ULaya.
 hata ki-iⁿgereza ni m-t̂jaⁿgaⁿ-ik-o w-a ki-latini na luya
 even DIM-English be VIT-mix-STAT-NOUN VIT-of DIM-Latin and language(N)[pl]
 j-iⁿgine z-a u-laja
 N.PL-other N.PL-of ABST-Europe
 ‘Even English is a mixture of Latin and other languages of Europe.’
- (99) KiSwahili hutuletea faida nyingi.
 ki-swahili hu-tu-let-e-a faida j-iⁿgi
 DIM-Swahili HAB-1PL-bring-APPL-IND advantage(N)[pl] N.PL-many
 ‘Swahili brings us many advantages.’
- (100) Husaidia WaAfrika wasiojua KiIngereza kusikizana na WaZungu.
 hu-saidia wa-afrika wa-si-o-fua ki-iⁿgereza
 HAB-help AN.PL-Africa AN.PL-NEG-REL-known DIM-English
 ku-siki-z-an-a na wa-zuⁿgu
 INF-hear-CAUS-RECP-IND with AN.PL-European
 ‘It helps Africans who do not know English to understand and be understood by Europeans.’
- (101) Husaidia WaAfrika wa makabila mbalimbali kusikizana.
 hu-saidia wa-afrika w-a ma-kabila ⁿbaliⁿbali ku-siki-z-an-a
 HAB-help AN.PL-Africa AN.PL-of AUG.PL-tribe different INF-hear-CAUS-RECP-IND
 ‘It helps Africans of different tribes to understand each other.’
- (102) Husaidia WaAfrika wanaokijua kusoma magazeti na matangazo ya serikali.
 hu-saidia wa-afrika wa-na-o-ki-fu-a ku-som-a
 HAB-help AN.PL-Africa AN.PL-PRS-REL-DIM-know-IND INF-read-IND
 ma-gazeti na ma-taⁿgaz-o j-a serikali
 AUG.PL-newspaper and AUG.PL-proclaim-NOUN AUG.PL-of government(N)
 ‘It helps those Africans who know it to read newspapers and government proclamations.’
- (103) Vitabu vilivyopigwa chapa katika KiSwahili ni vingi sana na vyenye maana kabisa katika maendeleo yetu.
 vi-tabu vi-li-vj-o-pig-w-a t̂japa katika ki-swahili
 DIM.PL-book DIM.PL-PST-DIM.PL-REL-hit-PASS-IND mark in DIM-Swahili
 ni v-iⁿgi sana na vj-eje maana kabisa katika
 be DIM.PL-many very and DIM.PL-having meaning(N) absolutely in
 ma-eⁿdele-o j-etu
 AUG.PL-progress-NOUN AUG.PL-our
 ‘Books that have been printed in Swahili are very numerous, and this fact has a great significance for our progress.’

- (104) KiSwahili kitaendelea na kitazidi kusitawishwa mpaka kiwe chema zaidi ya jinsi kilivyo sasa.

ki-swahili ki-ta-eⁿdele-a na ki-ta-zidi
 DIM-Swahili DIM-FUT-continue-IND and DIM-FUT-increase

ku-sitawi-ŷ-w-a mpaka ki-we tŷ-ema zaidi
 INF-prosper-CAUS-PASS-IND until DIM-be(SBJV) DIM-good more

ja finsi ki-li-vj-o sasa
 than way(N) DIM-be-DIM.PL-REL now

‘Swahili will expand and continue to flourish until it is even better than it is now.’
 [mpaka kiwe]. In a relative clause introduced by [mpaka] ‘until’, the verb has to be in the subjunctive mood.

[finsi ki-li-vj-o]. Subordinate clauses of manner or degree are introduced by [finsi] ‘manner’, and the verb complex in such clauses must contain, among other things, the relative suffix with diminutive plural agreement. The diminutive plural prefix [vi-] is often used to derive manner adverbials from adjective stems.

5.8 Exercises

5.8.1 SWAHILI

Analyze the following Swahili passage (Perrott 1951: 150, who cited *Maendeleo*) in the same style as our sample texts. We give the passage in standard Swahili orthography; you fill in the IPA pronunciation, the morpheme-by-morpheme gloss, and the free translation. We have supplied a few hints, but the rest of the information you need can be found in the Sketch of Swahili (§5.7).

- (1) Malezi Mabaya

ma-lezi ma-baya
 ABST.PL-upbringing ABST.PL-bad
 ‘Bad Upbringing’

- (2) Mara kwa mara tunawaona watoto wengi ambao hawana heshima na adabu nzuri kwa wakubwa wao,

? ? ? ? ? w-eⁿgi aⁿbao ?
 time(N) to ? subject-tense-object-see-mood ?-? gender.number-many ?-? ?-subject-be
 ? ? n-zuri ? ? wa-o
 respect(N) ? manners(N) gender-good ? ?-superior ?-their
 ‘From time ? who don’t have ?’

- (3) au pengine kwa wageni wanaofika katika mji fulani.

? ? ? ? ? ? ?
 or sometimes ? ?-foreigner ?-?-REL-arrive-? ? ?-city(VIT) certain
 ‘?’

fulani. Not all determiners take agreement markers.

- (4) Nimeona katika miji, hasa Nairobi, Mombasa na Dar es Salaam,
 ? ? ? hasa ? ? ? dar.es.salam
 ?-?-?-? ? ?-? especially ? ? ? ?
 ‘?’
- (5) mama wengine wenye watoto ambao hawawapeleki shuleni kusoma.
 ? ? ? ? ? ? ? ? ?
 mother(*gender.number*) ?-some ?-having ?-? ?-? ?-?-?-send-? ?-LOC ?-?-?
 ‘?’
- (6) Basi, mama hawa wakimwona mgeni, humwambia mtoto:
 ? ? ?-? ?-?-?-?-? ?-? ?-?-?-?-? ?-?
 then ?(*intrinsic gender*) DEM-*gender.number* ?-?-?-?-? ?-? ?-?-say-APPL-? ?-?
 ‘?’
- (7) “Mwombe peni ukanunue mkate.”
 ?-o^b-? ? ?-?-?-? ?-?
 ?-beg-mood ?(N) ?-?-buy-? ?-bread(VIT)
 ‘?’
- The subjunctive mood sometimes has imperative force.
- (8) Mtoto huondoka na kumwamkia yule mtu shikamuu kubwa,
 ?-? ?-?-? ? ?-?-?-?-? ?-? ?-? ? ?
 ?-? ?-set.off-? ? INF-?-greet-APPL-? AN-DEM ?-? greeting(N) big
 ‘?’
- na kumwamkia*. When two actions are closely associated, the verbs expressing these actions may be joined by [na] (comitative), and the second verb is then prefixed by [ku-] (infinitive).
- [shikamuu]. A humble greeting toward a social superior.
- (9) halafu akamwomba peni la mkate.
 ? ?-?-?-?-? ? I-a ? ?-?
 then ?-?-?-?-? ?(?) AUG-POSS ? ?-?
 ‘? penny for ?’
- (10) Peni lile mtoto hunyang’anywa na mama
 ? li-le ?-? ?-naŋaŋ-?-? ? ?
 ? *gender-DEM* ?-? ?-rob-?-? ? ?
 ‘?’

- (11) akanunua sigareti ama kitu kingine apendacho yeye.
 ?-?-?-? ? ? ?-? ?-? ?-?-[-]-? ?
 ?-?-?-? ? or ?-something ?-other ?-like-?-REL 3SG.AN
 ‘?’

5.8.2 ANALYSIS OF JU’HOAN

Examine the Ju’hoan sentences in (1) and their English glosses (from Snyman 1970) and then answer these questions:

- What signals grammatical relations in this language?
 - What is the basic word order?
 - How is number marked on nouns?
- [mi meni i!a] ‘I answer you.’
 - [i!a !aro mi] ‘You teach me.’
 - [!eŋ ho mi] ‘The eland sees me.’
 - [!h^hwā ho !eŋ] ‘The man sees the eland.’
 - [da?ama ho !eŋ] ‘The child sees the eland.’
 - [!h^hei ho !eŋ] ‘The lion/lions see(s) the eland.’
 - [de?ebi ho !eŋ] ‘The children see the eland.’
 - [!ae ho !eŋ] ‘The men see the eland.’
 - [!eu ho mi] ‘The elder sees me.’
 - [!eusi ho mi] ‘The elders see me.’

5.8.3 JU’HOAN PRONOUNS

Analyze the Ju’hoan data below, and on the basis of your analysis attempt to explain the differences in the pronouns. We give first a complete sentence, followed by a sentence in which each noun is replaced by a personal pronoun. For example, in (a), [ha ho ha] means ‘It (the child) sees him (the man)’.

- [da?ama ho !h^hwā]; [ha ho ha]. ‘The child sees the man; it sees him’
- [de?ebi ho !eusi]; [si ho si] ‘The children see the elders; they see them.’
- [!h^hwā wi dZ^fau]; [ha wi ha] ‘The man helps the woman; he helps her.’
- [eija [χ]oma de?ebi]; [ha [χ]oma si] ‘Mother pities the children; she pities them.’
- [wara ho !ae-kx^hao]; [ha ho ha] ‘The baboon sees the hunter; it sees him.’
- [wara ho eijasi]; [ho ho si] ‘The baboons see the mothers; they see them.’
- [!h^hao kwa^f |wara]; [ha kwa^f hi] ‘The badger fears the baboons; it fears them.’
- [!h^hao kwa^f |wara]; [hi kwa^f hi] ‘The badgers fear the baboons; they fear them.’

- i. [ʔ⁶ei-|eri kwa⁵ daʔa]; [ha kwa⁵ hi] ‘The wasp fears fire; it fears it.’
 j. [ʔ⁶ei-|eri kwa⁵ daʔasi]; [hi kwa⁵ hi] ‘The wasps fear fires; they fear them’
 k. [dʒ⁶au o ʔχ⁶anu]; [ha ho hi] ‘The woman sees the book; she sees it.’
 l. [dʒ⁶au o ʔχ⁶anusu]; [ha ho hi] ‘The woman sees the books; she sees them.’
 m. [dʒ⁶au ʔ²ama xore]; [ha ʔ²ama hi] ‘The woman buys the belt; she buys it.’
 n. [dʒ⁶ausi ʔ²ama xoresi]; [si ʔ²ama hi] ‘The women buy the belts; they buy them.’

5.8.4 LOANWORDS IN SWAHILI

Swahili has borrowed very heavily from both Arabic and English. In doing so it has adapted the original words to fit into the phonological patterns of Swahili. Examine the data below (Polomé 1967) and determine what general adaptations have taken place. Pay special attention to the final vowels in the Swahili forms.

5.8.4.1 Loanwords from Arabic

[adaʔu]	< [ʔadab]	‘good manners’
[kisi]	< [qis]	‘estimate’
[ratibu]	< [rat:ib]	‘arrange’
[wakati]	< [waqt]	‘time’
[madini]	< [maʔdin]	‘metal’
[maki]	< [maʔq]	‘thickness’
[milki]	< [milk]	‘possession, dominion’
[kaburi]	< [qabr]	‘grave, tomb’
[ðaifu]	< [d ⁵ aʔi:f]	‘weak’
[farifi]	< [far:iðʔ]	‘comfort’
[ʃahamu]	< [ʃahm]	‘fat, lard’
[duni]	< [du:n]	‘inferior, low’
[sakifu]	< [saq:if]	‘make a stone floor’
[kuzi]	< [ku:z]	‘earthenware pitcher’

5.8.4.2 Loanwords from English

[ʃuraʃi]		‘brush’
[ʃula ⁿ geti]		‘blanket’
[reli]	< <i>rail</i>	‘railway’
[eroplenu]		‘aeroplane’
[skrubu]		‘screw’
[paipu]	< <i>pipe</i>	‘motor horn’
[stimu]	< <i>steam</i>	‘power (of electricity)’
[kilabu]		‘club’
[madifaʔadi]		‘mudguard’
[steʃeni]		‘station’

5.9 Suggested readings

5.9.1 GENERAL

- ✘ *African languages: An introduction* (Heine & Nurse 2000). This collection's chapters offer an authoritative survey of the language families and their history.
- ✘ *African voices: An introduction to the languages and linguistics of Africa* (Webb & Kembo-Sure 2000). This textbook has more information than Childs (2003) on the sociolinguistics and politics of language in Africa.
- ✘ *Archaeology, language and the African past* (Blench 2006). Uses linguistic and archeological evidence to reconstruct the history of Africa and its languages; includes a 90-page introduction to African languages.
- ✘ *An introduction to African languages* (Childs 2003). Introductory level survey of African languages, including phonology, morphology, syntax, and classification.
- ✘ *A linguistic geography of Africa* (Heine & Nurse 2008). This book offers a typological approach to African languages, suggesting ways in which today's languages have built on past contacts.

5.9.2 AFRO-ASIATIC

- ✘ *Semitic languages: Outline of a comparative grammar* (Lipiński 2001). Comparative phonology, morphology, and syntax of Semitic languages. Part One, nearly 100 pages long, sketches the classical and modern languages and situates Semitic within Afro-Asiatic.

5.9.3 NILO-SAHARAN

- ✘ *The Nilo-Saharan languages* (Bender 1996). A scholarly comparison of over 600 lexical items shared among languages of this branch, testifying to cohesion within the branch despite its recognized diversity.

5.9.4 NIGER-CONGO

- ✘ *An introduction to African linguistics* (Mutaka 2000). Though the title doesn't suggest it, this book focuses on the structure of Bantu languages.
- ✘ *The Niger-Congo languages* (Bendor-Samuel & Hartell 1989). An overview of this family through detailed surveys by experts in its branches and subbranches.

5.9.5 KHOISAN

- ✘ Khoisan (Güldemann & Vossen 2000). Adopting the majority view that this is not a unified family, the authors divide Khoisan into five groups and examine the features of each.

5.9.6 ARABIC

- ✘ *Modern Arabic: Structures, functions, and varieties* (Holes 2004). A thorough yet readable, well-organized linguistic treatment of Arabic, with ample detail about the different spoken dialects.

5.9.7 SWAHILI

- ✘ *Modern Swahili grammar* (Mohamed 2001). An up-to-date introduction to Swahili phonology, morphology, and syntax.

Oceania

Geographically speaking, Oceania includes Australia and the majority of the island territories lying in the central and southern Pacific and Indian Oceans (Figure 6.1). There are three major language groups in the region: Austronesian, Australian, and Papuan languages. Only the first of these is widely recognized to be a language family (labeled “1” in Figure 6.1). The languages of Australia have many typological similarities to each other, but no sets of sound correspondences covering all the languages have been widely accepted. Papuan is a catch-all geographical grouping for languages in the region that are not in the Austronesian family and are not spoken on the Australian mainland; the term is not meant to imply that they are all genetically related to each other. It should be mentioned here, however, that Greenberg (1971) proposed that all Papuan languages, the extinct Tasmanian, and the Andamanese languages constitute a language family that he called *Indo-Pacific*. Although at least some major Papuan language groups are genetically related to each other, very few linguists now believe that Greenberg’s Indo-Pacific hypothesis as a whole is valid.

6.1 Austronesian

The name *Austronesian* means ‘southern’ (Latin *austr-*) ‘islands’ (Greek νῆσος): an appropriate name, because almost all the languages are spoken on islands of the south seas. The language family extends across the entire geographical region mentioned in the introduction to this chapter; see the languages labeled with a “1” in Figure 6.1. The total number of speakers of languages belonging to this family approaches 400 million, and it is rivaled only by Niger-Congo in the number of individual languages: the current Ethnologue counts 1,257 Austronesian languages. Despite its impressive size, there is no doubt that the languages are genetically related to each other (Table 6.1).



FIGURE 6.1 Some Oceanian languages. Families: ¹Austronesian. ²Pama-Nyungan.

TABLE 6.1

Some Austronesian languages

Name	Size	Location
Amis (AMI)	5	Taiwan
Paiwan (PWN)	5	Taiwan
Tsou (TSU)	4	Taiwan
Atayalic	6	Taiwan
· Atayal (TAY)	5	Taiwan
· Seediq (TRV)	5	Taiwan
Chamorro (CHA)	5	Guam
Palau (PAU)	5	Palau
Javanese (JAV)	8	Indonesia
Moken (MWT)	4	Burma
Philippine		Philippines
· Ilocano (ILO)	7	Philippines
· Tagalog (TGL)	8	Philippines
· Cebuano (CEB)	8	Philippines
· Maranao (MRW)	6	Philippines
Malayo-Sumbawan		
· Malay (MSA)	8	Indonesia
· Madurese (MAD)	8	Indonesia
· Sundanese (SUN)	8	Indonesia
· Balinese (BAN)	7	Indonesia
· Cham (CJA)	6	Cambodia
Greater Barito		
· Malagasy (MLG)	8	Madagascar
Central-Eastern		
· Tetum (TET)	6	East Timor
Oceanic		
· Yapese (YAP)	4	Micronesia
· Hiri Motu (HMO)	6	Papua New Guinea
· Micronesian		
· · Gilbertese (GIL)	5	Kiribati
· · Trukese (CHK)	5	Micronesia
· · Marshallese (MAH)	5	Marshall Islands
· Polynesian		
· · Tongan (TON)	6	Tonga
· · Samoan (SMO)	6	Samoa
· · Eastern Polynesian		
· · · Maori (MRI)	5	New Zealand
· · · Tahitian (TAH)	6	Tahiti
· · · Hawaiian (HAW)	4	Hawaii, U.S.
· · · Rapa Nui (RAP)	4	Easter Island, Chile

6.1.1.1 FORMOSA AREA

Most of the linguistic diversity within Austronesian is concentrated on the island of Taiwan, which was named *Formosa* ‘beautiful’ by Portuguese explorers. The original inhabitants of the island in historical times were speakers of Austronesian languages. In the seventeenth century AD, the island was colonized by the Dutch and subsequently by the Chinese. Nowadays, Taiwan is the locus of the Republic of China, and at least 98% of the population are ethnic Chinese. A useful if arbitrary convention is to use the word *Taiwanese* in reference to the various Chinese languages spoken on the island, and to refer to the aboriginal Austronesian languages as *Formosan*. There is no known genetic connection between the two sets of languages.

Formosa presents a challenge to genetic subgrouping, because the languages spoken there have existed side by side for thousands of years in a relatively small space. They have, therefore, influenced each other so heavily that it is difficult to separate the shared innovations due to shared intermediate protolanguages from the shared innovations due to diffusion across genetic boundaries. All researchers agree that many of the Formosan languages are very different from each other and that they do not themselves form a proper genetic clade sharing a set of innovations unique to the island. Rather, the term *Formosan* is a geographical term of convenience. The standard analysis is that of Blust (1999), who identified nine primary branches of Austronesian on Formosa, a selection of which is given at the top of Table 6.1.

6.1.1.2 MALAYO-POLYNESIAN

Ninety-eight percent of the Austronesian languages belong to the Malayo-Polynesian branch; the remainder are spoken in Taiwan. Malayo-Polynesian has many subbranches, here presented roughly in geographic order.

6.1.2.1 Micronesia area

In Micronesia, Chamorro and Palau are each difficult to classify, and at the moment are best treated as independent branches of Malayo-Polynesian.

6.1.2.2 Philippine

Most of the languages of the Philippines constitute their own, Philippine, branch of Malayo-Polynesian. Tagalog is the national language of the Republic of the Philippines and is widely understood by non-native speakers in that country. The form used as a standardized, national language is known as *Filipino*. Other widely spoken Philippine languages include Cebuano, Ilocano, Hiligaynon (HIL), and Waray-Waray (WAR).

6.1.2.3 Indonesia area

There is a great deal of diversity among the Malayo-Polynesian languages of the Indonesia area, which we take to include Malaysia, Singapore, Brunei, and East

Timor; some of the language groups also spill over into the Philippines and parts of the Asian mainland.

By far the dominant branch in this area is Malayo-Sumbawan. It includes Malay, a family of closely related languages or dialects. The national language of Indonesia, Bahasa Indonesia (IND), is a form of Malay, as is Bahasa Malaysia, an official language in Malaysia and Singapore. The institution of these national languages makes Malay, especially in its form Bahasa Indonesia, the most widely used of all Austronesian languages, if mostly by second-language speakers. Several other Malayo-Sumbawan languages have many speakers, but are giving way to those national languages. These include Madurese, Sundanese, and Balinese. Malayo-Sumbawan is one of the few branches of Austronesian that is spoken on a mainland. Part of Malaysia is located on the Asian mainland south of Thailand. Cham is spoken in Cambodia and Vietnam.

Some other Malayo-Polynesian languages of Indonesia do not belong to the dominant Malayo-Sumbawan group. Working from west to east, Sumatra, the homeland of the Malayo-Sumbawan languages, is also the main locus of four smaller branches of Malayo-Polynesian: Enggano (ENO), Lampung, Northwest Sumatra-Barrier Islands, and Rejang (REJ). A Javanese branch is found on Java. If only native speakers of a language are counted, Javanese has the greatest number of speakers of all the Austronesian languages. Borneo is the main location of Land Dayak, a North Borneo group, and a Greater Barito group. This last group also includes Malagasy, which is spoken far to the west in Madagascar. Finally, Sulawesi, which was formerly known as *Celebes*, hosts a Celebic group as well as a set of South Sulawesi languages, whose best-known member is Buginese (BUG).

The Moklen (MKM), or Moken, people, who mainly live in boats along the coasts of Thailand and Burma, speak a set of closely related languages that are generally considered their own branch of Malayo-Polynesian.

Languages of the Indonesian area have the earliest written records of the Austronesian family. The earliest inscriptions in Old Cham and Old Javanese date to the ninth century AD, and Old Malay goes as far back as 683 AD. All of these languages were written in scripts derived from those used in India.

6.1.2.4 Central-Eastern Malayo-Polynesian

Most of the Malayo-Polynesian languages are members of a large clade called *Central-Eastern Malayo-Polynesian*. This clade has hundreds of languages, but only a few million speakers altogether.

CENTRAL MALAYO-POLYNESIAN AREA

Although there is some evidence that the Central Malayo-Polynesian group forms a genetic clade, the more conservative approach is to treat it as a geographical grouping. It includes some 100 languages spoken in the eastern parts of Indonesia, especially in the Maluku (Molucca) Islands and the Lesser Sunda Islands; a few are spoken in western New Guinea. The largest language in this grouping may be Tetum, an official language of East Timor.

SOUTH HALMAHERA–WEST NEW GUINEA

There is more evidence for a South Halmahera–West New Guinea clade. This includes about 41 smaller languages spoken for the most part in coastal Irian Jaya (Indonesia) and coastal New Guinea.

OCEANIC

The Oceanic subbranch of Central-Eastern Malayo-Polynesian is a well-established clade. It contains over 500 languages, or about half of the total for the Austronesian family as a whole. Most Oceanic languages have lost word-final consonants and have simplified consonant clusters. Many of the languages, including Hawaiian, have no codas and consonant clusters at all.

The branches and subbranches of Oceanic are too numerous to list here in full. Generally speaking, a number of very small subbranches are found in coastal New Guinea and adjacent islands, including the Admiralty Islands of the Bismarck Sea, the Solomon Islands, Vanuatu, and the Loyalty Islands. Oceanic also includes the languages of Micronesia except for the previously discussed Palauan and Chamorro; in a special Micronesian clade are Gilbertese (Kiribati), Trukese, and Marshallese. Yapese appears to be an Oceanic language, but has proven difficult to classify.

Perhaps the best-known group within Oceanic is Central Pacific. It includes a clade that comprises Fijian (FIJ) and Rotuman (RTM).

Central Pacific also includes the Polynesian languages, which also form a proper clade. Among them are Tongan and Samoan, the latter spoken in Samoa and various places in the United States, including American Samoa. A particularly close-knit subgroup are the Eastern Polynesian languages, which include Maori, Tahitian, Hawaiian, and Rapa Nui, among others. Maori and Hawaiian surrendered many speakers to English in the course of the twentieth century, but both New Zealand and Hawaii now have many enthusiasts dedicated to preserving the use of those languages.

6.1.3 GENERAL FEATURES OF AUSTRONESIAN

It is very difficult to make general statements about Austronesian languages, because they are spread over such a large geographical area and are typologically quite diverse. The basic word order for Austronesian is predominantly VOS on Taiwan, in the Philippines, and in adjacent parts of Borneo and Sulawesi, as well as on Madagascar. It is SVO in much of Indonesia and Melanesia. A number of the Austronesian languages of New Guinea are SOV, as a result of contact with Papuan languages. Polynesian languages are VSO.

As was already stated in §4.5, Paul Benedict considers Austronesian languages to belong to his Austro-Tai language family, along with Tai-Kadai languages, Hmong-Mien languages, and Japanese. Other scholars have attempted to link Austronesian with Altaic, Indo-European, Austroasiatic (the Austric hypothesis),

and Sino-Tibetan, but so far none of these hypotheses has been thoroughly explored and confirmed by solid evidence.

A great deal of interest has been devoted to the question of where the original homeland of the Proto-Austronesian language family was located. An original homeland – the more concise German term *URHEIMAT* is also used – is the place where a language clade's protolanguage was spoken. There are two different but complementary approaches that may be taken in order to determine where the *urheimat* of any protolanguage was located.

One possibility is to employ the principle that the area of greatest linguistic diversity within a given language family must be the area that was settled earliest by the speakers of languages belonging to the language family in question, and that areas of little linguistic diversity must be the areas of more recent settlement. This principle is derived from observing the linguistic situation in the areas for which we have ample historical documentation. For example, it is clear that the most pronounced differentiation among the various dialects of American English is found on the Atlantic seaboard, which was settled earliest by European colonists; the farther west one goes, the less dialect differentiation can be found. In China, too, the spread of Mandarin into southwestern China and northeastern China has been relatively recent, accounting for the fact that there is still much similarity between Mandarin dialects spoken in Sichuan and Yunnan (southwestern China) and the dialects spoken in northeastern China (former Manchuria). On the other hand, on the coast of China, areas settled by Chinese-speaking peoples much earlier, there are sometimes very deep linguistic divisions even among adjacent speech communities.

The other basis for situating the original homeland involves reconstructing the vocabulary of the protolanguage and then examining this vocabulary for clues as to the material culture of the speakers of the protolanguage, including terms for flora and fauna. Thus, it is argued that if the protolanguage had the word for 'coconut', its speakers must have lived in a tropical area where coconuts were found. In addition, archeologists may attempt to match up various prehistoric archeological sites with a particular reconstructed protolanguage, based on the presence at the archeological sites of artifacts, animals, evidence of cultivation of various crops, and so forth, all of which match the presence in the reconstructed protolanguage of words for such items. Unfortunately, this strategy is not without its pitfalls, because words for various artifacts, unfamiliar plants and animals, may diffuse from language to language, and, if borrowed before the major sound changes occurred in the borrowing languages, may no longer be recognizable as loanwords and thus be reconstructed as part of the original vocabulary.

Blust (1985) and Bellwood (1991) attempted to apply both of the foregoing approaches in determining the homeland of Austronesian. From a purely linguistic point of view, Proto-Austronesian was spoken on the island of Taiwan, perhaps 5,000 years ago. Even today, the greatest linguistic diversity is found on that island; all branches of Austronesian except for Malayo-Polynesian are found

on Taiwan. Bellwood has shown that the speakers of Proto-Austronesian must have come originally from coastal south China, where, according to archeological evidence, there were Neolithic settlements whose inhabitants developed rice and sugar cane cultivation, domesticated water buffalo and other cattle, and had such implements as plows, axes, and canoes. However, there are no traces remaining of any early Austronesian language in China. As the Austronesian language speakers expanded in various directions, this expansion resulted in further splits from the family tree, and as the speakers developed new artifacts or encountered new plants and animals in their new habitats, lexical items for these new things were added to their languages. Thus, for example, lexical items for things such as taro, breadfruit, yam, banana, sago, and coconut cannot be reconstructed for Proto-Austronesian, but they can be reconstructed for Proto-Malayo-Polynesian, which was likely spoken in an area south of the Proto-Austronesian homeland, perhaps the Celebes Sea.

6.2 Papuan area

As already stated at the beginning of this chapter, *Papuan* is a term of convenience that covers all languages of the Pacific that do not belong to Austronesian. By *Pacific*, we mean in practice an area stretching from eastern Indonesia and East Timor to the Solomon Islands. This area explicitly excludes the mainland of Australia, but does include the Torres Strait. The Papuan languages are those in Figure 6.1 that do not have a prefixed “1” and are not located in Australia. The center of the Papuan world is clearly New Guinea, where most of the Papuan languages are spoken and where most of the languages are Papuan.

The Papuan languages have been very resistant to cladistic analysis. Sober assessments count about 800 languages forming dozens of different families, with quite a few language isolates. Although there have been various proposals to link these families together, most of them remain speculative. A large part of the problem is that the great majority of Papuan languages are very poorly documented. A recent attempt at a large-scale analysis ended up being based primarily on forms of personal pronouns, because so little is known about most of the languages (Ross 2005). Unfortunately, pronominal forms tend to be short and often resemble those of other languages by chance. More problematic still is the issue discussed earlier for the languages of Formosa: loans. The languages have coexisted for millennia side-by-side, most of them on the same island (New Guinea), in a state of high multilingualism. Under such circumstances, one would not be surprised to see languages borrow not only vocabulary but many grammatical features as well.

There have been attempts over the years to bring this situation under control by proposing larger groupings. Greenberg (1971) proposed that all the Papuan languages were related to each other, and also to Tasmanian, the languages of the Andaman Islands, and Kusunda (KGG), a language spoken in Nepal. This proposal

has been roundly rejected by specialists. On a narrower scale, he broke down the Papuan languages proper into five groups. But even these less ambitious groupings did not match very closely the groupings that other specialists have proposed.

A proposal that has received much more support among specialists posits a large language family called *Trans–New Guinea*, so called because the languages are spoken in various locations all *across* the length of the island of New Guinea. Many specialists in Papuan linguistics accept the proposal in its broadest terms, but differ in which languages they include. At its height, as reflected in the 15th edition of *Ethnologue*, *Trans–New Guinea* was listed as having 564 languages. Newer assessments, especially that of Ross (2005), have reduced that number, so that the 16th edition listed only 477 languages. One of the more concrete claims made for *Trans–New Guinea* languages is that they share an agglutinative morphology in which verbs take a prefix that marks the person of the object: [n] for first person and [g] for second person, to which is added [a] for the singular and [i] for the plural. Although this is an impressively specific finding, it is difficult to assess whether it is enough to prove that the languages are related, or whether it might be coincidental. Very short morphemes, especially inflectional affixes, may turn up in different languages by chance more often than one might expect, and the likelihood of a match between languages increases dramatically if one allows exceptions, such as a language that has first person [n] but not second person [g]. Furthermore, even pronoun affixes may be borrowed when languages are in close contact for long periods of time. These observations are not meant to belittle the work of historical linguists specializing in Papuan languages, but to increase our appreciation of how difficult their task is, and of why different researchers often come to substantially different interpretations of the facts.

If we limit discussion only to clades of languages that are provably related to each other, there are perhaps five dozen families, far too many to usefully list. Nor would it be very helpful to list the over 800 individual Papuan languages. Such sources as *Ethnologue* provide reasonably comprehensive listings. Some of the more widely spoken Papuan languages, by country, are shown in Table 6.2.

If this meager selection seems stingy, it is only fair to point out that the national governments of the area are at a similar loss when it comes to foregrounding

TABLE 6.2

Some Papuan languages

Name	Size	Location
Western Dani (DNW)	6	Indonesia
Enga (ENQ)	7	Papua New Guinea
Makasae (MKZ)	5	East Timor
Meriam (ULK)	3	Australia
Bilua (BLB)	4	Solomon Islands

the Papuan languages. In most of the countries that host Papuan languages, it is Austronesian languages that tend to dominate both officially and in terms of unofficial use as a second language. In Indonesia, Bahasa Indonesia is the official language; in East Timor, Tetum is the indigenous official language; and in Papua New Guinea, the indigenous official language is Hiri Motu; all three are Austronesian. Except in Indonesia, however, even stronger competition now comes from European languages: Portuguese in East Timor and English in Papua New Guinea and the Solomon Islands, though for the most part it is actually a creole based on English vocabulary that is more widely used than international standard English (more on creoles in §8.4). Although it may seem odd for countries to embrace the language of external subjugating powers, they often find that the sting is diminished greatly with the achievement of independence. When a country has many ethnic groups, each with its own language, an external world language can often come across as a relatively neutral choice for a language of government and inter-ethnic communication.

According to Foley (1992), Papuan languages tend to be agglutinating in type and tend to be mainly suffixing. They also tend to have SOV as their basic word order, with modifiers preceding their head nouns and with postpositions instead of prepositions. There has been some diffusion of linguistic features from Austronesian languages into neighboring Papuan languages, and vice versa.

6.3 Australian area

According to its 2011 census, Australia is an overwhelmingly English-speaking country, with substantial minorities speaking such languages as Chinese, Italian, Arabic, and Greek (Australian Bureau of Statistics 2013). All these are European and Asian languages, introduced into Australia in or after 1788. Prior to European colonization, Australia was host to hundreds of Aboriginal languages. Many of them still exist, but the vast majority are dead or moribund, with the children and child-bearing adults being fluent only in English, virtually guaranteeing that future generations will not be native speakers of Aboriginal languages. The National Indigenous Languages Survey reported:

The situation of Australia's languages is very grave and requires urgent action. Of an original number of over 250 known Australian Indigenous languages, only about 145 Indigenous languages are still spoken and the vast majority of these, about 110, are in the severely and critically endangered categories. This critically endangered category indicates languages that are spoken only by small groups of people mostly over 40 years old. Eighteen languages are strong in the sense of being spoken by all age groups, but three or four of these are showing some signs of moving into endangerment. (Australian Institute of Aboriginal and Torres Strait Islander Studies 2005: 3).

See Figure 6.1 for the location of some representative languages. The common convention is to apply the term *Australian* to Aboriginal languages spoken in what is now the country of Australia. However, one of the two languages of the Torres Strait Islands, Meriam (ULK), is grouped with the Papuan languages even though it is spoken in Queensland. Occasionally one hears claims that the original languages of Tasmania should not properly be counted among the Australian languages, but the only meaningful support for that viewpoint was the Indo-Pacific hypothesis that held that Tasmanian languages were related to the Papuan and Andamanese languages (Greenberg 1971), a hypothesis that is now roundly rejected. In fact, all Aboriginal Tasmanian languages were exterminated so early that there is little data to support any sort of genetic analysis.

Table 6.3 includes the largest Australian languages and a handful of representative smaller ones. Be aware that Australian language names tend to be spelled in a bewildering variety of ways. The first Australian language to be recorded in writing has been spelled as Gogo-Yimidjir, Gugu-Yimidhirr, Gugu Yimijir,

TABLE 6.3

Some Australian languages

Name	Size	Location
Anindilyakwa (AOI)	4	Northern Territory
Murrinh-Patha (MWF)	4	Northern Territory
Tiwi (TIW)	4	Northern Territory
Gunwingguan		
· Gunwinggu (GUP)	4	Northern Territory
Bunuban		
· Gooniyandi (GNI)	2	Western Australia
Jarrakan		
· Kija (GIA)	2	Western Australia
Nyulnyulan		
· Nyigina (NYH)	2	Western Australia
Pama-Nyungan		
· Dyrbal (DBL)	1	Queensland
· Guugu Yimidhirr (KKY)	3	Queensland
· Kala Lagaw Ya (MWP)	3	Queensland
· Dhuwal (DUJ)	4	Northern Territory
· Warlpiri (WBP)	4	Northern Territory
· Pitjantjatjara (PJT)	4	South Australia
· Arandic		
· · Eastern Arrernte (AER)	4	Northern Territory
· · Western Arrernte (ARE)	4	Northern Territory
· · Alyawarr (ALY)	4	Northern Territory
· · Anmatyerre (AMX)	4	Northern Territory

Guguyimidjir, Guugu Yimidhirr, Guugu Yimithirr, Koko Imudji, Koko Yimidir, Kuku Jimidir, Kuku Yimidhirr, Kuku Yimithirr, and Kukuyimidir. Such variation highlights the comparative convenience and uniformity of the ISO 639-3 codes, which is KKY for this language.

6.3.1 PAMA-NYUNGAN

About two thirds of the languages of Australia belong to a single family, Pama-Nyungan. Their territory comprises about 90% of the continent. As can be seen in Figure 6.1, Pama-Nyungan languages, which are labeled with a “2”, are spoken everywhere except in a relatively small band in the north of the continent. The name shows the great geographical extent of the family. It is compounded from the word for ‘man’ at two extremes of the Pama-Nyungan area: *pama* in the northeast of Australia (the Cape York peninsula of Queensland) and *nyunga* [ɲuŋa] in the southwest. But the great majority of the most populous Pama-Nyungan languages are spoken nowadays in the Northern Territory. These include Dhuwal and Warlpiri. A particularly important subgroup of Pama-Nyungan is the Arandic clade, which includes such populous languages as Eastern Arrernte, Western Arrernte, Alyawarr, and Anmatyerre, which are spoken in and around Alice Springs. Various dialects of Western Desert, including Pitjantjatjara, are also found in South Australia and Western Australia.

The Torres Strait lies between New Guinea and the Australian mainland. This is Australian territory, but the Torres Strait Islanders consider themselves ethnically Papuan. One of the two languages spoken in the islands is the mixed language Kala Lagaw Ya. It is a lingua franca that has borrowed extremely heavily from Austronesian and Papuan languages, but it appears that its core grammar, including its pronouns, is Pama-Nyungan.

6.3.2 NON-PAMA-NYUNGAN AREA

In the north, stretching from West Australia to Queensland, is a relatively small area where the bulk of Australia’s linguistic diversity is located. Here are found many small language families and not a few language isolates. The non-Pama-Nyungan languages are those that appear on the map of Australia without a prefixed number (Figure 6.1). Perhaps the most diverse of these non-Pama-Nyungan families, with some 25 languages, is Gunwingguan, named after its largest language, Gunwinggu. Some very small families are the Nyulnyulan family, represented by Nyigina; Bunuban, which includes Gooniyandi; and Jarrakan, which includes Kija. Quite a few language isolates are also found in the north; among them are Tiwi, Anindilyakwa, and Murrinh-Patha. Tiwi is notable for being a polysynthetic language in which nouns can be incorporated within verbs. For example, Osborne (1974) reports a word that he glossed as ‘He came and stole my wild honey this morning while I was asleep’. However, it has been reported that in

the current generation, this complexity has been lost, probably due to intensive influence by English.

6.3.3 LANGUAGE CHARACTERISTICS

Although the languages of Australia cannot, in our present state of knowledge, be grouped into a single family, typologically they are very similar. Much of this similarity may be due to extensive borrowing.

The sound systems are very uniform. There is an almost total absence of fricatives and affricates, a feature that is unusual outside of Oceania. In the vast majority of Australian languages, voicing is not lexically contrastive. They do have voiced and voiceless stops, such as [t] and [d], but these are almost always allophones of the same phoneme. Most languages have a phonemic contrast between alveolar and retroflex consonants, such as [t] versus [ɽ], both sounds made with the tip of the tongue. This contrast extends to rhotic consonants. The alveolar rhotic is [r], a TRILL or TAP familiar from languages like Spanish; the other is [ɽ], an approximant familiar from some dialects of English. Australian languages also have laminal stops and often laminal laterals; LAMINAL CONSONANTS are produced with the flat of the blade of the tongue. Alveolopalatal laminals are found everywhere, and many languages also have dental laminals. The great majority of Australian languages have three or four different coronal places of articulation, which is unusually high and reminiscent of the inventories of some Dravidian languages. The vowel systems, in contrast, are usually small, the triangular [i], [u], and [a].

Most Australian languages have split-ergative alignment. For example, the Western Desert language has nominative-accusative case marking in pronouns but has ergative-absolutive case marking in nouns. Dyirbal has a similar pattern; see §6.5.6.2 in the Sketch of Dyirbal for a fuller discussion of this property. The usual word order is typically reported as being SOV, and adjectives tend to follow the nouns they modify. But speakers can and do place the components of a sentence in any order they choose, and the different orders seem to have little if any impact on the meaning of the sentence. That is, the word order is very free, so much so that there may be no real evidence that any particular order is more basic or that clauses are built from phrases rather than individual words. Hale (1981) called attention to these facts in Warlpiri, initiating a vigorous discussion of the implications such syntax has for mainstream syntactic theory in the United States, which treats phrase structure as fundamental to the syntax of all languages.

Because they developed in relative isolation from the rest of the world until very recently, Australian languages have often been exploited by anthropologists and psycholinguists interested in exploring LINGUISTIC RELATIVITY: the WHORFIAN HYPOTHESIS that distinctive properties of specific languages make speakers of those languages think differently. Haviland (1993) called attention to the fact that in Guugu Yimidhirr, positions are exclusively described in

terms of points of the compass (north, east, etc.) rather than in terms relative to the interlocutors, such as the speaker's left or right. Various observations and experiments showed that speakers of Guugu Yimithirr not only developed an exceptional sense of and memory for geographic orientation, but also categorized objects differently from Europeans. In one experiment (Levinson 1997), a participant was positioned facing north and was shown a card with a blue chip on the left and a red chip on the right. Then he was taken into another room, positioned so that he was facing south, and asked to pick the exact same card. Whereas speakers of Dutch in this situation picked a card with a blue chip to the left of the red chip, the great majority of Guugu Yimithirr speakers did the opposite. From their point of view, a card would be classified as identical to the first one they saw only if it had a blue chip to the west of the red chip. Because the task was nonverbal, researchers concluded that the ways positions are expressed in Guugu Yimithirr and Dutch fundamentally affect the way speakers of those languages look at the world.

6.4 Sketch of Hawaiian

This account of the structure of Hawaiian is based chiefly on that found in Elbert & Pukui (1979), with some additional analyses and examples by Greg Lee Carter.

6.4.1 GENETIC AFFILIATION AND GENERAL INFORMATION

Hawaiian belongs to the Polynesian subgroup of the Oceanic branch of Austronesian, along with such languages as Maori, Samoan, Tahitian, and a host of other languages spoken on various islands in the Pacific Ocean, especially in the South Pacific and Central Pacific within the Polynesian triangle bounded by Hawaii in the north, New Zealand in the south, and Easter Island (Chile) in the east.

Like most Polynesian languages, Hawaiian has relatively simple phonology. It is typologically an analytic language, though it does have some affixation.

Although perhaps 2,000 people with Hawaiian ancestry still speak Hawaiian, almost all the speakers are bilingual in English (or Hawaii Creole) and many do not speak the language on a daily basis anymore. Only on the privately owned island of Ni'ihau, off the coast of Kaua'i, is there a small community of about 100 Hawaiian speakers for whom Hawaiian is truly a native language.

In recent years there has been a revival of interest in Hawaiian language and culture among the people living in Hawaii; as a result, the Hawaiian language is studied more widely in schools, and spoken Hawaiian is heard more often on the radio, joining the sung Hawaiian that has always been an important part of Hawaiian music.

6.4.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

Hawaiian is known for having one of the smallest consonant inventories in the world, though it does exceed the six-consonant inventory attributed to Rotokas (ROO), a Papuan language. Table 6.4 lists the consonant phonemes.

The paucity of the consonant phonemes is due to a number of historical MERGERS of Proto-Polynesian consonants in Hawaiian. A merger occurs when one or more consonants change into a sound that already existed in the language. For example, both *[s] and *[f] have merged into existing [h].

[w] is pronounced either as [v] or [w]. These two sounds are in FREE VARIATION in Hawaiian; that is, speakers may choose to use either variant. For more details, see Schütz (1980).

[k] is usually pronounced [t] on the island of Ni‘ihau except at the beginning of a sentence. On other islands, it is usually pronounced [k]. But there are words that seem to be pronounced with a [t] everywhere. For example, /ku:ku:/ ‘grandma’ or ‘grandpa’ is usually pronounced as [tu:tu:] by all speakers of Hawaiian.

In comparison to the inventory of consonants, the vowel inventory is fairly rich, consisting of the classical five vowel system, which is further enhanced by a length contrast for all vowels (see Table 6.5).

In addition to the monophthongs, there are also the following short diphthong phonemes:

- (1) iu, ei, eu, ai, ae, ao, au, oi, ou

and the following long diphthong phonemes:

- (2) ai:, a:u, a:e, a:o, ei:, ou:

All other combinations of vowels act as sequences of separate vowel phonemes, in separate syllables.

As Schütz (1980: 25) pointed out, if diphthongs are considered to be individual phonemes, the total number of phonemes adds up to 33. This is certainly not such a small number of phonemes to deserve special mention as being unusual. However, the number of consonant phonemes is relatively low by any count.

TABLE 6.4

Hawaiian consonant phonemes

Stop	p	k	ʔ ʻ
Fricative			h
Nasal	m	n	
Lateral		l	
Glide	w		

TABLE 6.5

Vowel phonemes of Hawaiian

High	i	i:		u	u:
Mid	e	e:		o	o:
Low			a	a:	

Primary stress is placed on the final syllable if it is long (contains a diphthong or long vowel); if the final syllable is short, the primary stress is placed on the penultimate syllable.

Hawaiian spelling agrees very closely with the IPA, with the exception of glottal stops and vowel length. Prior to World War II, most Hawaiian texts ignored those features, even though they are phonemic. Nowadays, the glottal stop is officially marked by a reversed apostrophe, and long vowels are marked with a MACRON: a horizontal bar over the vowel, as in ⟨ā⟩.

Hawaiian has borrowed extensively from various European languages, especially English. It also has several loanwords from Southern Chinese languages like Yuè and Mǐnnán, because Chinese laborers from South China formed the first wave of indentured workers that came to Hawaii to work on the sugar cane plantations.

- (3) a. [pipi] ‘beef’
 b. [ka:la:] ‘dollar’
 c. [pele] ‘bell’
 d. [kalikimaka] ‘Christmas’
 e. [konohi:] ‘Chinese New Year’ < Mǐnnán [kɔŋʰiʷ] ‘congratulations’

Because Hawaiian does not allow any consonant clusters or coda consonants, Hawaiian versions of borrowed words have vowels inserted between consonants or at the end of the word after a consonant. Another way of dealing with consonant clusters in borrowed words is to drop some of the consonants altogether. Also, since Hawaiian has no lingual fricatives or any affricates, usually [k] is substituted for English [s], [ʃ], [tʃ], [dʒ], and the like, making some of the loanwords from English almost unrecognizable.

6.4.3 MORPHOLOGY AND SYNTAX

6.4.3.1 Noun phrase

A typical noun phrase is composed of the following elements, in a fairly fixed word order:

- (4) determiner + plural + noun + modifier

For example:

- (5) keia mau hale nui
 DEM.PROX PL house big
 ‘these big houses’

Modifiers that are adjectives in English translation are actually stative verbs – verbs that show a state rather than an action. A word like [nui], for example, basically means ‘be big’ (compare 43).

NOUN

Hawaiian is basically an analytic language and does not have much inflection. However, there are 10 nouns, all of them referring to people, including some kinship terms, which have special plural forms, formed by lengthening the antepenultimate vowel.

- (6) a. [kahuna] ‘priest’, [ka:huna] ‘priests’
 b. [kupuna] ‘grandparent’, [ku:puna] ‘parents’
 c. [wahine] ‘woman’, [wa:hine] ‘women’

Plurality is more commonly marked by preceding the noun with the marker [mau] or the plural determiner [na:].

DETERMINERS

When a common noun is head of a noun phrase, it is normal for it to be preceded by a determiner. The determiners comprise articles, demonstratives, and possessives.

ARTICLES

The singular definite article is [ke] or [ka] (DEF.ART.SG) (7a–c). The decision of which allomorph to use is mostly determined by the phonetics. The [ke] allomorph is used before words beginning with the vowels [a], [e], or [o], or with [k] (7a). Elsewhere, [ka] is used (7b–c). But there are some words that take [ke] even though they begin with [ʔ] or [p] (7d). The plural definite article is [na:] (7e–f). There is also an indefinite article, [he] ‘a, some’ (7g), which is most common in verbless sentences. Like many grammatical markers in Hawaiian, the articles are proclitic.

- (7) a. [ke=aliʔi] ‘the chief’
 b. [ka=?aka] ‘the laugh’
 c. [ka=iʔa] ‘the fish’
 d. [ke=?ala] ‘the fragrance’
 e. [na:=aliʔi] ‘the chiefs’
 f. [na:=wa:hine] ‘the women’
 g. [he=aliʔi] ‘a chief’
 h. [he=mau aliʔi] ‘chiefs’

The indefinite determiner [ke=kahi] ‘a certain’ often introduces a new item into the discourse.

DEMONSTRATIVES

Hawaiian has a large number of demonstratives. Those that most commonly occur in the determiner position are the following. These three words combine the base [ke:-] with another morpheme that itself can be used as a demonstrative. For example, [la] is used enclitically as a mild demonstrative.

- (8) a. [ke:-ia] proximal (DEM.PROX) ‘this’ (near speaker)
 b. [ke:-na:] MEDIAL ‘that’ (near addressee)
 c. [ke:-la:] DISTAL (DEM.DIST) ‘that’ (away from both the speaker or addressee)

All of these can also stand alone without a noun. The next most common demonstrative, [nei], usually follows the noun it depends on.

GENITIVE

Genitive relations like possession are introduced by the proclitics [a], [a:], [o], and [o:]. The distinction between the forms with short and long vowels is phonological: the long vowel is used if the proclitic attaches to a word whose first syllable has a long vowel or diphthong. The distinction between the two vowels [o] and [a] is semantically conditioned, signaling the nature of the relationship between the possessor and what is possessed. Thus, [a] indicates that possessor caused or had some active control over the ownership, whereas [o] indicates that the possessor had no control over the ownership. Some of the terms used to describe this contrast are ALIENABLE POSSESSION (ALIEN) versus *inalienable possession* (INAL), *acquired* versus *inherited*, and *active* versus *passive*. This contrast is signaled by alternating the vowels [a] and [o] in all possessive words. However, there are other considerations that override the basic rule as given above. One such rule is that location expressions always use [o]. Thus, for example, clothing one is wearing or a horse one is riding involves [o]-marked possession, regardless of who owns the clothing or the horse.

- (9) a. ke=ki?i a=pua
 DEF.ART.SG=picture ALIEN=Pua
 ‘Pua’s picture (which she created or owns)’
 b. ke=ki?i o=pua
 DEF.ART.SG=picture INAL=Pua
 ‘Pua’s picture (e.g. a photograph of her)’
 c. ke=ali?i a=kalani
 DEF.ART.SG=chief ALIEN=Kalani
 ‘Kalani’s chief (whom he appointed)’
 d. ke=ali?i o ka=?a:ina
 DEF.ART.SG=chief INAL DEF.ART.SG=land
 ‘the (hereditary) chief of the land’

TABLE 6.6

Hawaiian personal pronouns

Person	SG	DU	PL
1.EXCL	au	ma:-ua	ma:-kou
1.INCL		ka:-ua	ka:-kou
2	ʔoe	ʔo-lua	ʔou-kou
3	ia	la:-ua	la:-kou

Possessive phrases can follow the head noun, as in the above examples, or they can precede it. When they precede the noun, the possessive proclitic fuses with the definite article, and the possessor noun phrase fills the determiner slot:

- (10) a. ke=keiki a=pua
 DEF.ART.SG=child ALIEN=Pua
 ‘the child of Pua’
- b. k-a:=pua keiki
 DEF-ALIEN=Pua child
 ‘Pua’s child’

PRONOUNS

The pronominal system of Hawaiian is relatively complicated (Table 6.6). There are dual forms as well as singular and plural. The words for ‘we’ also distinguish between an inclusive and exclusive meaning, as we saw in Běijīng Mandarin.

The first person singular adds a glottal stop when not used as a nominative: [aʔu]. The dual number suffix is derived from [lua] ‘two’, and the plural marker is derived from [kolu] ‘three’. There are no gender distinctions. Thus, [ia] can mean ‘he’, ‘she’, and even ‘it’, although it is often omitted when referring to inanimate objects.

The possessive pronouns consist basically of one of the possessive proclitics [a:] or [o:] attaching to the pronoun, in the same way they can attach to nouns. Thus ‘your and mine’ (i.e. inclusive dual) is [a:=ka:-ua] or [o:=ka:-ua], depending on whether the possession is alienable or inalienable.

- (11) a. ke:-ia keiki a:=ka:-ua
 DEM-PROX child ALIEN=1.INCL-DU
 ‘this child of ours’

The singular possessive forms are irregular:

- (12) Person INAL ALIEN
 1 a-ʔu o-ʔu
 2 a:-u o-u
 3 a:-na o-na

We have seen how demonstratives, when used as definite determiners before nouns, usually fuse with the definite article. Possessives do the same thing, except that the article drops its vowel before the possessive.

- (13) a. [k-a:=ka:-ua keiki] ‘your and my child’
 b. [k-o-ʔu kupuna] ‘my grandparent’
 c. [k-a:-na wahine] ‘his wife’
 d. [k-o-na hale] ‘his house’

VERBLESS SENTENCES

Hawaiian makes extensive use of verbless sentences, obviating the need for verbs like English *be* and *have*. A juxtaposition of two noun phrases either identifies the general category to which the second belongs (14a) or equates the two. Locative elements are often juxtaposed with a noun phrase to tell where or when it occurs (14b). Juxtaposing a noun phrase with a possessive is the usual way of showing possession (14c). BENEFACTIVES (BEN) in Hawaiian are formed by attaching [n] to the beginning of the basic possessive form. These forms are typically found at the beginning of a verbless sentence. They are so called because they may show who some thing or action is for, but often their meaning is indistinguishable from general possession (14d–e).

- (14) a. he=lawaiʔa au
 INDF=fisherman 1SG
 ‘I am a fisherman.’
- b. aia he=lu:ʔau i=ka=po:-ʔa-lima
 there INDF=feast LOC=DEF.ART=evening-number-five
 ‘There’s a feast Friday.’
- c. he=hale nui k-o-na
 INDF=house big DEF-INAL-3SG
 ‘He has a big house.’
- d. n-o=ka=lani ka=inoa
 BEN-INAL=DEF.ART.SG=royal DEF.ART.SG=name_chant
 ‘The name chant is about the royal chief.’
- e. n-a=ke=keiki na:=hulu
 BEN-ALIEN=DEF=child DEF.ART.PL=feather
 ‘The feathers belong to the child.’

6.4.3.2 Verb morphology

As you have already seen, in Hawaiian there are many proclitics that mark various grammatical relations. Affixation is limited on nouns and is not much more common on verbs. Below are listed some of the more important verbal prefixes.

[hoʔo-]

The addition of this prefix usually causes a verb to become a DELIBERATE transitive verb. A deliberate verb is one that typically has an animate agent who does the action on purpose.

- (15) a. [hele] ‘go’ → [hoʔo-hele] ‘to set in motion’
 b. [makaʔu] ‘to fear’ → [hoʔo-makaʔu] ‘to frighten’

Especially before nouns, this prefix has a SIMULATIVE function (‘be like’):

- (16) [haole] ‘white person’ → [hoʔo-haole] ‘to act like a white person’

[ʔo:-]

This prefix adds the meaning of ‘somewhat’:

- (17) [maʔi] ‘sick’ → [ʔo:-maʔi] ‘sickly’

[aka-]

This signifies ‘carefully, slowly’:

- (18) [ʔai] ‘to eat’ → [aka-ʔai] ‘to eat slowly’

6.4.3.3 Verb phrase structure

The basic order of elements in a clause is VSO.

CASE PROCLITICS

Grammatical relations between the verb and its dependents are marked by placing a proclitic particle before the noun phrases (Table 6.7). As we did for Tibetan, we will call these *case* particles, even though they are not expressed by inflection.

The nominative marker is used for the subject of a sentence, if it is definite. It tends to be omitted when the subject immediately follows a verb, especially if it has some other marker of definiteness. Thus it is more common before proper names (19a) and some pronouns, especially third-person singular [ia] (19b). And it normally appears on definite subjects in verbless sentences, especially if they begin the sentence (19c).

TABLE 6.7

Case proclitics in Hawaiian

Proclitic	Function
ʔo, Ø	subject (NOM)
i, ia:	object (ACC), location (LOC)
e	agent of passive (AGT)
na	fronted agent (AGT)
e	address (VOC)
ma(:)	location (LOC)
mai	'from' (ABL)
me	'with' (COM)
na, no	'for' (BEN)

- (19) a. hauʔoli ʔo=pa:pa:
 be_happy NOM=Papa
 'Papa is happy.'
- b. ua ʔike ʔo=ia
 PFV see NOM=3SG
 'He saw (it).'
- c. ʔo=ka=lei pu:pu: ka=lei o=niʔihau
 NOM=DEF.ART.SG=garland shell DEF.ART.SG=garland INAL=Ni'ihau
 'The shell garland is the garland of Ni'ihau.'

The accusative proclitic marks an object of the verb. The variant [ia:] occurs before all pronouns and proper names. Elsewhere, the form [i] is used.

- (20) ha:ʔawi ke=kanaka i=ka=makana ia:=pua
 give DEF.ART.SG=person[nom] ACC=DEF.ART.SG=present ACC=Pua
 'The man gives the present to Pua.'

Direct objects and indirect objects both take the accusative marker, but can be distinguished because the direct object comes first.

The same proclitic can also mark location. It is traditional to gloss it as locative in such contexts, in part because the accusative and the locative proclitics came from two different etymological sources.

- (21) ua=noho ke=kanaka i=hilo
 PFV=stay DEF.ART.SG=person LOC=Hilo
 'The man stayed in Hilo.'

In this use, its meaning is scarcely distinguishable from that of [ma] (pronounced [ma:] when the next vowel is long or diphthongal).

If the verb is passive – followed by the passive enclitic [ʔia] – the agent may be expressed with [e]:

- (22) ua=lohe=ʔia ka=moa e=kekahi po:poki
 PFV=hear=PASS DEF.ART=chicken AGT=some cat
 ‘The chicken was heard by a certain cat.’

If the agent is fronted to the beginning of the sentence, before the verb, it is preceded by [na]. In such an event, the object, which is now directly after the verb, often appears without the object marker [i].

- (23) na=pua i=ka:kau (i=)ka=leka
 AGT=Pua PFV=write (ACC=)DEF.ART=letter
 ‘It was Pua who wrote the letter.’

The VOCATIVE proclitic is used to mark a person or thing being called or addressed.

- (24) e=pua hele mai
 VOC=Pua go hither
 ‘Pua, come here!’

Use of the other case particles is fairly straightforward. It is important to bear in mind, though, that, like English prepositions, a small number of particles are put to a great many uses that are difficult to explain.

VERB MARKERS

Verb markers are particles that signal tense, aspect, and mood. In Hawaiian, they come immediately before and after the verb.

The marker [ua] signals finished or completed action or a change of state. Thus it is a marker of perfective aspect. A variant form used when other elements precede the verb in the sentence is [i].

- (25) a. ua=hele ke=kanaka
 PFV=go DEF.ART.SG=person
 ‘The man has gone.’
 b. ʔaʔole i=hele ke=kanaka
 NEG PFV=go DEF.ART.SG=person
 ‘The man hasn’t gone.’

When a noun phrase precedes the verb, the particle [ai] must also be added after the perfective verb.

The imperfective aspect may be signaled by placing [e] before the verb. The verb is then usually followed by [ana].

- (26) e=hele aku ana ke=kanaka
 IPFV=swim away IPFV DEF.ART.SG=person
 ‘The man is going away.’

When the verb follows a noun phrase, the particle [ai] replaces [ana].

The present progressive, which appears as such only at the beginning of a sentence, is signaled by placing [ke] before the verb and a demonstrative, usually [nei], after it.

- (27) ke=kali nei au
 PRS.PROG=wait PRS.PROG 1SG
 ‘I am waiting.’

The imperative mood is signaled by preceding the verb with [e] (28a). [o:] is a more polite alternant, and a more informal alternative is to use no marker at all. The subject ‘you’ is included much more often than in English. In addition to a command, these markers signal intent, purpose, necessity, and probability – the sort of things associated with a separate subjunctive in other languages (28b). A related function of [e] is to subordinate a verb to the main verb of a sentence (28c–d).

- (28) a. e=hele ?oe
 IMP=go 2SG
 ‘Go!’ or ‘You ought to go!’
- b. e=?au?au ?o=kawelo
 IMP=bathe NOM=Kawelo
 ‘Kawelo should bathe.’
- c. hele aku=la ?o=kawelo e=?au?au
 go away=DEM NOM=Kawelo SUBORD=bathe
 ‘Kawelo went to bathe.’
- d. makemake au e=?au?au
 want 1SG SUBORD=bathe
 ‘I want to bathe.’

The negative imperative, or prohibitive, is formed with the marker [mai]. This marker is used only with the second person.

- (29) mai=ue: ?oe
 PROH=cry 2SG
 ‘Don’t cry.’

After two common verbs, [hiki] ‘be possible’ and [pono] ‘right, should, must’, [ke] is used instead of [e].

- (30) hiki ia=?u ke=?ai
 be_able ACC=1SG SUBORD=eat
 ‘I can eat.’

[hiki], by the way, is a member of a small set of verbs that appear to swap the subject and object roles, at least from an English speaker’s perspective. From a Hawaiian perspective, it may be better to think of such verbs as stative verbs whose subject is the thing characterized by the state and whose object is the cause of the state: ‘(It) is possible for me to eat’. Another such verb is [loa?a] ‘find’, ‘get’:

- (31) loa?a ka=?aihue ia=?u
 catch DEF.ART=thief[nom] ACC=1SG
 ‘I have caught the thief.’

DIRECTIONALS

There are four different directional particles that are frequently placed after the verb.

[mai] indicates movement toward the speaker and is therefore often translated as ‘to me’ or ‘here’. This is known as a *VENITIVE* directional, from Italian *venire* ‘come’. Its function is very much like the German prefix *her-*.

- (32) hele mai
 go VEN
 ‘Come!’

With verbs of speaking, the venitive indicates a reply, though not necessarily to the speaker.

[aku] signals movement away from the speaker, like German *hin-*. This is known as an *ANDATIVE* (AND) directional, from Italian *andare* ‘go’.

- (33) hele aku
 go AND
 ‘Go away!’

With verbs of speaking, this directional indicates that someone else is being addressed. In other words, the speaker is not simply speaking to himself.

[iho] signals motion downward (*DOWN*) as well as reflexive action.

- (34) a. hele iho
 go down
 'Go down!'
- b. ke=na:na: nei au ia=?u iho
 PRS.PROG=look_at PRS.PROG 1SG ACC=1SG DOWN
 'I am looking at myself.'

[a?e] signals movement upward (UP) or such notions as adjacency in space or time.

- (35) hele a?e
 go UP
 'Go up!'

NOMINALIZERS

The most common nominalizer is the enclitic [?ana] (36) and the suffixes [-na], [-hana], [-kana], and [-lana].

- (36) ka=hele ?a:wi:wi=?ana mai
 DEF.ART.SG=go quick=NOUN VEN
 'the quick arrival'

The different consonants that appear in the alternants of the suffixes for this marker, as well as with the passive and in transitivizing suffixes, were originally parts of the verb stems, which were reanalyzed as being part of the suffix.

The decision as to whether nominalized verbs take [a] (alienable) or [o] (inalienable) possessives depends on whether the subject normally has deliberate control over the activity.

- (37) a. [k-a:-na ?ai=?ana] 'his eating'
 b. [k-o-na aloha] 'his love'

PASSIVE MARKER

The most general passivizing marker is the enclitic [?ia], which has suffixal alternants like those of nominalizers.

- (38) ua=?a:pono wale=?ia ke=ka:nawai
 PFV=approve unnecessarily=PASS DEF.ART.SG=law
 'The law was approved unnecessarily.'

TRANSITIVIZERS

There are a number of suffixes that have the shape [-Caʔi] and [-Ci] (for different consonants C) and that make transitive verbs out of intransitive verbs. However, in many cases they form new verbs with new meanings that are not predictable:

- (39) [pi:] ‘to sprinkle’ → [pi:-naʔi] ‘to fill a crack’

REDUPLICATION

Verb bases may be reduplicated either fully or partially. Reduplication signifies frequentative or otherwise increased action, perhaps by multiple agents. Such functions are subsumed under the term AUGMENTATIVE.

- (40) [ni:nau] ‘to question’ → [nina-ninau] ‘to question repeatedly; to interrogate; for many people to ask questions’

6.4.3.4 Subordinate clauses

There are several ways that English temporal subordinate clauses may be rendered in Hawaiian. One way is by using the locative proclitic [i]~[ia:] in a temporal sense, to mean ‘when’ or ‘while’:

- (41) ia=?u e=noho ana me=?ou-kou
 LOC=1SG IPFV=stay IPFV COM=2-PL
 ‘while I was staying with you’

Another way is with a nominalized verb phrase and a locative proclitic:

- (42) i=k-o-na hele=?ana mai
 LOC=DEF-INAL-3SG go=NOUN VEN
 ‘at his coming’, ‘while he was coming’

They may also be expressed by words meaning ‘time’, ‘epoch’, ‘era’, and so on, which can have modifiers:

- (43) i=k-o-ʔu wa: liʔiliʔi nui k-o=ma:-kou leʔaleʔa
 LOC=DEF-INAL-1SG age little big DEF-INAL=1.EXCL-PL fun
 ‘When I was little, we had a lot of fun.’
 Literally, ‘In my childhood age, our fun was great.’

There is also a conjunction [ke], which means ‘when’ and introduces subordinate clauses that refer to the future, but only in declarative statements, not questions:

- (44) ke=po:loli ʔai
 when=hungry eat
 ‘When hungry, eat.’

Hawaiian relative clauses have no relative pronoun: they contain no element corresponding to English *who*, *which*, or *that*. To convert a sentence like (45a) so that the subject is modified by a relative clause (45b), the subject noun phrase is placed first, followed by the verb and the rest of the elements. The only adjustment made within the relative clause itself is that some of the tense and aspect markers change; in this example, the perfective marker [ua] became [i]. If some element other than a subject is to be modified, however, like the direct object ‘book’ in this example, it is rare to treat the subject of the relative phrase as its grammatical subject. Instead, the subject is pulled out and converted to a possessive phrase. Thus, although (45c) seems to begin ‘the book of the woman’, [a=ka=wahine] is actually the subject of the relative clause. The embedded verb of such relative clauses must also be followed by an aspect marker; [ai] is supplied for aspects that would not otherwise have an aspect marker after the verb.

- (45) a. ua=ka:kau ka=wahine i=ka=puke
 PFV=write DEF.ART=WOMAN ACC=DEF.ART=book
 ‘The woman wrote the book.’
- b. ka=wahine i=ka:kau i=ka=puke
 DEF.ART=WOMAN PFV=write ACC=DEF.ART=book
 ‘the woman who wrote the book’
- c. ka=puke a=ka=wahine i=ka:kau ai
 DEF.ART=book INAL=DEF.ART=WOMAN PFV=write PFV
 ‘the book that the woman wrote’

6.4.3.5 Interrogative sentences

Polar questions may be distinguished from statements only by intonation. Thus (46), as transcribed without intonation, can be either declarative or interrogative.

- (46) ?ike ?oe ia:=keawe
 know 2SG ACC=Keawe
 ‘Do you know Keawe?’

The main interrogative words used in content questions are [aha] ‘what’, [hea] ‘where’ or ‘which’, [wai] ‘who’, and [-hia], a numeral interrogative.

- (47) a. he=aha ke:=la:
 INDF.ART=what DEF.DEM=DIST
 ‘What is that?’
- b. ?o=wai ke=kumu
 NOM=who DEF.ART.SG=teacher
 ‘Who is the teacher?’

- c. ʔe-hia ka:lɑ:
 number-how_many dollar
 ‘How much money?’

6.4.4 SAMPLE TEXT

This little folk tale is taken from Hopkins (1992: 190). It may well be that this tale is not only Hawaiian but has wider Polynesian roots. Author Lyovin recalls first hearing this story from a Tongan student at the University of Hawaii, who represented it as a Tongan folk tale.

- (48) Ka He'e a me ka 'Iole

ka=heʔe a=me=ka=ʔiole
 DEF.ART.SG=octopus and=COM=DEF.ART.SG=rat

‘The Octopus and the Rat’

[a me]. The word [a] can mean ‘and’ by itself, but is usually extended by [me], literally ‘with,’ before a noun phrase.

- (49) I ka wā kahiko ua noho kekahi 'iole me kona 'ohana ma ka moku-puni 'o Mokoli'i.

i=ka=wa: kahiko ua=noho kekahi ʔiole
 LOC=DEF.ART.SG=time ancient PFV=live a_certain rat

me=k-o-na ʔohana ma=ka=mokupuni ʔo=moko-liʔi
 COM=DEF-INAL-3SG family LOC=DEF.ART.SG=island NOM=Lizard-Little

‘In ancient times there lived a rat with his family on Mokoli'i island.’

- (50) (I kēia manawa, ua kapa 'ia kēia wahi 'o Chinaman's Hat.)

i=ke:=ia manawa ua=kapa=ʔia ke:=ia wahi
 LOC=DEF.DEM=PROX time PFV=call=PASS DEF.DEM=PROX place

ʔo=tʃanəmənzhæt

NOM=Chinaman's Hat

‘(Today this place is called *Chinaman's Hat*.)’

- (51) I kekahi lā, makemake 'o ia e hele aku i Kāne'ohe.

i=kekahi la: makemake ʔo=ia e=hele aku i=ka:neʔohe
 LOC=a_certain day want NOM=3SG SUBORD=go AND LOC=Kāne'ohe

‘One day, he wanted to go to Kāne'ohe.’

- (52) Ua loa'a iā ia he wa'a, a hoe akula 'o ia i Kualoa.

ua=loaʔa ia:=ia he=waʔa a=hoe aku=la ʔo=ia i=kualoa
 PFV=get ACC=3SG INDF.ART=canoe and=paddle AND=DEM NOM=3SG LOC=Kualoa

‘He got a canoe and paddled over to Kualoa.’

- (53) I kona hiki ‘ana mai, ua ho‘opa‘a ‘ia ka wa‘a ma ke kumu niu, a hele akula ka ‘iole i Kāne‘ohe.
 i=k-o-na hiki=?ana mai ua=hoʻo-paʻa=?ia ka=waʻa
 LOC=DEF-INAL-3SG arrive=NOUN VEN PFV=CAUS-fast=PASS DEF.ART.SG=canoe
 ma=ke=kumu niu a=hele aku=la ka=?iole i=ka:neʻohe
 LOC=DEF.ART.SG=trunk coconut and=go AND=DEM DEF.ART.SG=rat LOC=Kāne‘ohe
 ‘On his arrival, the canoe was tethered to the trunk of a coconut tree and the rat went off to Kāne‘ohe.’
- (54) Akā, ‘a‘ole i ho‘opa‘a pono ‘ia ka wa‘a.
 aka: ?aʻole i=hoʻo-paʻa pono=?ia ka=waʻa
 but NEG PFV=CAUS-fast well=PASS DEF.ART.SG=canoe
 ‘However, the canoe was not tethered well.’
- (55) No laila, i kona hiki ‘ana mai mai Kāne‘ohe mai, ua lilo ka wa‘a i ke kai.
 no=laila i=k-o-na hiki=?ana mai mai=ka:neʻohe mai
 for=there LOC=DEF-INAL-3SG arrive=NOUN VEN ABL=Kāne‘ohe VEN
 ua=lilo ka=waʻa i=ke=kai
 PFV=be_lost DEF.ART.SG=canoe LOC=DEF.ART.SG=ocean
 ‘Therefore, by the time the rat returned from Kāne‘ohe the canoe had been lost in the ocean.’
- (56) Hū ka pilikia nui!
 hu: ka=pilikia nui
 INTERJECTION DEF.ART.SG=trouble big
 ‘Oh, what a disaster!’
- (57) Ua noho ihola ka ‘iole a uē me ka leo nui loa.
 ua=noho iho=la ka=?iole a=ue: me=ka=leo nui loa
 PFV=sit DOWN=DEM DEF.ART.SG=rat and=lament COM=DEF.ART.SG=voice big very
 ‘The rat sat down and wept very loudly.’
- (58) Ua lohe ‘ia kona uē ‘ana e kekahi he‘e lokomaika‘i.
 ua=lohe=?ia k-o-na ue:=?ana
 PFV=hear=PASS DEF-INAL-3SG lament=NOUN
 e=kekahi heʻe loko-maikaʻi
 AGT=a_certain octopus character-good
 ‘His crying was heard by a kind octopus.’
- (59) Nīnau akula ka he‘e iā ia, “E ia nei, he aha kou pilikia?”
 ni:nau aku=la ka=heʻe ia:i=ia
 ask AND=DEM DEF.ART.SG=octopus ACC=3SG

e=ia nei he=aha k-o-u pilikia
 VOC=3SG here INDF.ART=what DEF-INAL-2SG trouble

‘The octopus asked him, “Hey buddy, what’s the matter?”’

[E ia nei] literally ‘O he here’ is an affectionate way of addressing someone.

- (60) A pane maila ka ‘iole, “Ua nalowale ko’u wa’a; ua lilo paha i ke kai, no ka mea ‘a’ole maika’i paha ko’u ho’opa’a ‘ana.

a=pane mai=la ka=?iole ua=nalowale k-o-?u wa?a
 and=answer VEN=DEM DEF.ART.SG=rat PFV=disappear DEF-INAL-1SG canoe

ua=lilo paha i=ke=kai no=ka=mea ?a?ole
 PFV=be_lost perhaps LOC=DEF.ART.SG=ocean for=DEF.ART.SG=thing NEG

maika?i paha k-o-?u ho?o-pa?a=?ana
 good perhaps DEF-INAL-1SG CAUS-fast=NOUN

‘The rat replied saying, “My canoe has disappeared; it may be lost in the ocean because I maybe did not fasten it well.’

[ko?u wa?a]. Canoes, like other conveyances, are regarded as inalienable regardless of their actual ownership.

- (61) ‘A’ole hiki ia’u ke ho’i aku i Mokoli’i no ka mea ‘a’ole hiki ia’u ke ‘au’au.’

?a?ole hiki ia=?u ke=ho?i aku i=moko-li?i
 NEG be_able ACC=1SG SUBORD=leave AND LOC=Lizard-Little

no=ka=mea ?a?ole hiki ia=?u ke=?au?au
 for=DEF.ART.SG=thing NEG be_able ACC=1SG SUBORD=swim

“Now it is not possible for me to go back to Mokoli’i because I can’t swim.”’

- (62) Pane maila ka he’e na’au palupalu, “A’ole pilikia; hiki paha ia’u ke kōkua aku iā ‘oe.

pane mai=la ka=he?e na?au palupalu
 answer VEN=DEM DEF.ART.SG=octopus heart tender

?a?ole pilikia hiki paha ia=?u ke=kōkua aku ia=?oe
 NEG trouble be_able perhaps ACC=1SG SUBORD=help AND ACC=2SG

‘The kind-hearted octopus then replied, “No problem; maybe I can help you out.’

- (63) E pi’i a’e ‘oe i luna o ku’u po’o, a e ho’iho’i aku au iā ‘oe i kou home.”

e=pi?i a?e ?oe i=luna o=k-u-?u po?o
 IMP=climb up 2SG LOC=top INAL=DEF-POSS-1SG head

a=e=ho?iho?i aku au ia=?oe i=k-o-u home
 and=IMP=return AND 1SG ACC=2SG LOC=DEF-INAL-2SG home

“Climb up on top of my head, and let me take you back to your home.”’

[ku?u po?o]. [ku?u] means ‘my’, usually applied to something or someone one has affection toward. It can be used regardless of whether the possession is alienable or not.

- (64) I ke kau ‘ana o ka ‘iole ma luna o kona po‘o, ua ho‘o-maka ka he‘e e ho‘i aku i Mokoli‘i.

i=ke=kau=?ana o=ka=?iole ma=luna
 LOC=DEF.ART.SG=perch=NOUN INAL=DEF.ART.SG=rat LOC=top
 o=k-o-na po?o ua=ho?o-maka ka=he?e
 INAL=DEF-INAL-3SG head PFV=CAUS-start DEF.ART.SG=octopus
 e=ho?i aku i=moko-li?i
 SUBORD=return AND LOC=Lizard-Little

‘When the rat mounted on top of his head, the octopus set out to return to Mokoli‘i.’

- (65) Hū ka nui o kō ka ‘iole maka‘u, akā, kāhea akula ka he‘e, “E noho mālie ‘oe!

hu: ka=nui o=k-o:=ka=?iole maka?u
 INTERJECTION DEF.ART.SG=big INAL=DEF-INAL=DEF.ART.SG=rat fear
 aka: ka:hea aku=la ka=he?e e=noho marlie ?oe
 however cry_out AND=DEM DEF.ART.SG=octopus IMP=sit calm 2SG
 ‘Oh, how the rat was afraid! But the octopus called out saying, “Stay calm!”

- (66) Mai maka‘u! Kama‘āina loa au i kēia kai.”

mai maka?u kama-?a:ina loa au i=ke:=ia kai
 PROH be_afraid child-land much 1SG LOC=DEF.DEM=PROX ocean
 ‘“Don’t be afraid! I am quite at home here in this sea.”’

- (67) Ma hope iho, ua hiki akula lāua i Mokoli‘i.

ma=hope iho ua=hiki aku=la la:-ua i=moko-li?i
 LOC=after DOWN PFV=arrive AND=DEM 3-DU LOC=Lizard-Little
 ‘Afterwards they reached Mokoli‘i.’

- (68) Ua iho ihola ka ‘iole mai luna iho o kō ka he‘e po‘o, a ha‘i akula ‘o ia iā ia, “Ke mahalo nui loa nei au iā ‘oe no kou lawe ‘ana mai ia‘u.

ua=iho iho=la ka=?iole mai=luna iho
 PFV=descend DOWN=DEM DEF.ART.SG=rat ABL=top DOWN
 o=k-o:=ka=he?e po?o a=ha?i aku=la ?o=ia
 INAL=DEF-INAL=DEF.ART.SG=octopus head and=say AND=DEM NOM=3SG
 ia:=ia ke mahalo nui loa nei au ia:=?oe
 ACC=3SG PRS.PROG thank big very PRS.PROG 1SG ACC=2SG
 n-o=k-o=u lawe=?ana mai ia=?u
 for-INAL=DEF-INAL=2SG bring=NOUN VEN ACC=1SG

‘The rat jumped down from the top of the octopus’s head and then said to him, “Thank you very much for bringing me here.’

- (69) “Aia kekahi makana nāu ma luna o kou po’o,” a holo ʔāwīwī akula ka ʔiole i uka.
 aia kekahi makana n-a:=u ma=luna o=k-o-u poʔo
 there a_certain present for-ALIEN=2SG LOC=top INAL=DEF-INAL-2SG head
 a=holo ʔa:wī:wī: aku=la ka=ʔiole i=uka
 and=run quickly AND=DEM DEF.ART.SG=rat LOC=inland
 “‘There is a present for you on top of your head.’ Then the rat ran off inland very quickly.”
- (70) Ua hāhā aʔela ka he’e ma luna o kona po’o, a loa’a ke kūkae.
 ua=ha:ha: aʔe=la ka=heʔe
 PFV=grope UP=DEM DEF.ART.SG=octopus
 ma=luna o=k-o-na poʔo a=loaʔa ke=ku:kāe
 LOC=top INAL=DEF-INAL-3SG head and=find DEF.ART.SG=excrement
 ‘The octopus groped on top of his head and found excrement.’
- (71) Ua ki’o ka ‘iole ma luna o ke po’o no ka mea maka’u loa ‘o ia.
 ua=kiʔo ka=ʔiole ma=luna o=ke=poʔo
 PFV=excrete DEF.ART.SG=rat LOC=top INAL=DEF.ART.SG=head
 n-o=ka=mea makaʔu loa ʔo=ia
 for-INAL=DEF.ART.SG=thing fear big NOM=3SG
 ‘The rat had defecated on top of the octopus’ head because he was greatly frightened.’
- (72) Hū ka huhū nui o ka he’e!
 hu: ka=huhu: nui o=ka=heʔe
 INTERJECTION DEF.ART.SG=anger great INAL=DEF.ART.SG=octopus
 ‘Oh, how angry the octopus was!’
- (73) Mai kēlā manawa mai, inā ‘ike ‘ia ka leho e ka he’e, ua lālau ‘o ia iā ia no ka mea,
 mana’o ‘o ia, ‘o ka ‘iole nō ia.
 mai ke:=la: manawa mai ina: ʔike=ʔia ka=leho
 from DEF.DEM=DIST time VEN if see=PASS DEF.ART.SG=cowry_shell
 e=ka=heʔe ua=la:lau ʔo=ia ia:=ia
 AGT=DEF.ART.SG=octopus PFV=grab NOM=3SG ACC=3SG
 n-o=ka=mea manaʔo ʔo=ia ʔo=ka=ʔiole no: ia
 for-INAL=DEF.ART.SG=thing think NOM=3SG NOM=DEF.ART.SG=rat indeed 3SG
 ‘From that time on, if an octopus sees a cowry shell, he grabs at it, because he thinks that it must be the rat.’
- (74) Hana ka po’e Pākīpika i kekahi mea i kapa ‘ia “lūhe’e” me ka pōhaku a me ka leho,
 a loa’a ka he’e iā lākou me kēia mea.
 hana ka=poʔe pa:ki:pika i=kekahi mea i=kapa=ʔia lu:heʔe
 make DEF.ART.SG=people Pacific ACC=a_certain thing PFV=call=PASS “lūhe’e”

me=ka=po:haku a=me=ka=leho
 COM=DEF.ART.SG=stone and=COM=DEF.ART.SG=cowry_shell
 a=loa?a ka=he?e ia:=la:=kou me=ke:=ia mea
 and=catch DEF.ART.SG=octopus ACC=3=PL COM=DEF.DEM=PROX thing

‘The people of the Pacific make a device called *lūhe’e* with a stone and a cowry shell and they catch octopuses with it.’

6.5 Sketch of Dyirbal

6.5.1 GENETIC RELATIONSHIP AND GENERAL BACKGROUND

Dyirbal (DBL; also spelled Jirrbal, Chirpalji, Djirbal, Djirubal, Dyirrbal, and Tjirbal) is spoken in northeastern Queensland, Australia. Today it has at most five speakers and is classified by UNESCO as critically endangered (Moseley 2010, language-id-178). Indeed, as long ago as 1985 it was the subject of a book on language death (Schmidt 1985), which highlighted the profound changes that took place in the language as its speakers increasingly shifted to dominant or exclusive use of English. In this sketch, we deal exclusively with the traditional language as described by Dixon (1972). Our sketch, including the examples, is based almost exclusively on that book, supplemented by Dixon (1994; 2011) and a few other sources.

Dyirbal belongs to the Pama-Nyungan family of languages. Dixon (1972) described three different dialects, but we discuss here only the main dialect, Jirrbal. When Dixon researched the language, there was no indigenous name for the language, just for the dialects, which were spoken by different tribes. Dixon named the language after the main dialect, but used a spelling variant to distinguish the language from the dialect. The names *Dyirbal* and *Jirrbal* are pronounced exactly the same.

6.5.2 MOTHER-IN-LAW SPEECH

A single language can have several registers. The same person may use different registers in different social settings. To a large extent these registers are distinguished by differences in vocabulary. In an informal discussion among close friends, for example, it may be perfectly appropriate to say “Don’t put up with that crap!” In more formal settings, one might rephrase the whole sentence to use more learned and less viscerally emotive vocabulary, perhaps resulting in “Toleration of such misconduct is not advised.”

Choice of register depends heavily on the audience and the speaker’s relationship with the audience – social relationships that differ from society to society. A particularly distinctive type of register in Aboriginal Australia was one that developed in order to express linguistically the special relation a person had with certain relatives. Among Dyirbal speakers, there were strong taboos in place

limiting contact between a man and his mother-in-law and between a woman and her father-in-law. There were also taboos in place between a man and his first cousin, but only if that first cousin was a woman who belonged to the section of the tribe from which he was allowed to select a wife. People were not supposed to approach tabooed relatives or look at them, much less speak with them. In Dyirbal society, there was a register, called *Jalnguy*, which had to be used in any conversation that was within earshot of a tabooed relative. One might also use the special register when talking to relatives who were the same-sex counterparts of tabooed relatives; for example, a man would usually use *Jalnguy* when speaking to his father-in-law, who was not tabooed but required respect as husband of the tabooed mother-in-law. Thus *Jalnguy* was an avoidance register in the sense that its function was to keep tabooed relatives at a social distance. Because the taboo between mother-in-law and son-in-law was especially strong throughout Australia, this register is commonly referred to as mother-in-law speech.

In most Australian languages, the mother-in-law register contained dozens of distinctive words that were not used in the common registers, but Dyirbal carried the pattern to its logical conclusion. With just a few exceptions – mostly kinship terms – the content words of *Jalnguy* were completely unique to that register. No content word from the common register could be used when speaking *Jalnguy*. Furthermore, the lexical words in *Jalnguy* bore no phonetic similarity to the corresponding lexical word in the common register: for example, the *Jalnguy* word for ‘look at’ was [ɲuɟimal], while the common word was [puɟal]. Despite these strict requirements, *Jalnguy* was a fully functional register, in which one could hold full conversations or tell stories. The difficulty of learning an entire parallel vocabulary was eased somewhat by the fact that *Jalnguy* relied heavily on highly general words. Thus while the common language has in addition to [puɟal] several different hyponyms (more specific terms) that mean ‘look up at’, ‘watch someone going’, ‘stare’, ‘look at with a light’, ‘spy on’, and so on, in *Jalnguy* one made do with just [ɲuɟimal]. One could always be more specific by using more expansive descriptions. In this way the *Jalnguy* vocabulary was kept at about a quarter the size of the common vocabulary.

Jalnguy died out of common use by around 1930 with the breakdown of the taboo system in Dyirbal society. In this sketch only the everyday Dyirbal register is described.

6.5.3 PHONETICS AND PHONOLOGY

6.5.3.1 Consonants

Dyirbal has a simpler consonant inventory than most other Australian languages in that it lacks dental and retroflex stops (Table 6.8). The alveolopalatal stops [ç] and [ɲ] are laminal, with the body of the tongue raised toward the hard palate. They can be thought of as heavily palatalized versions of the basic [t] and [n] sounds. The IPA has no specific symbols for these sounds, although one can get

TABLE 6.8

Consonant phonemes of Dyirbal

	Bilabial	Alveolar	Alveolopalatal	Retroflex	Velar
Stops	p	t	c		k
Nasals	m	n	ɲ		ŋ
Liquid lateral		l			
Liquid rhotic		r		ɻ	
Glides			j		w

fairly close by using diacritics such as [tʲ] or [ç]. In the interest of simplicity, we use the latter representation but drop the plus sign, which means that the sound is produced more forward in the mouth than a strictly palatal [ç] should be.

As in most Australian languages, there is no distinction between voiceless and voiced stops. They are all unaspirated stops for which voicing is allophonic. As a general rule, the stops are voiceless unless surrounded by voiced sounds, in which case they are voiced. We treat the voiceless allophones as the more basic, unmarked forms and the voiced ones as being the result of voicing assimilation; hence our choice of symbols for phonemic notations.

6.5.3.2 Vowels

Because Dyirbal has only three vowel phonemes (75), there is room for wide-ranging allophonic variation. For example, [a] can be realized as [ɛ] after alveolopalatals and as [ɔ] after velars.

(75) i u
 a

Stress is generally placed on the first syllable of a word and on every alternating syllable except the last. There are several complications with stress, but they are not dealt with here.

6.5.4 ORTHOGRAPHY

We know of no official orthography for Dyirbal. The most common approach is to spell the three vowels as in IPA but to write the consonants using English letter values. Thus [j] is spelled ⟨y⟩ and [ɻ] is spelled ⟨r⟩. The stops are spelled with the letters that suggest voiced sounds, such as ⟨b⟩, ⟨d⟩, ⟨j⟩, and ⟨g⟩, because the voicing of Dyirbal stops works almost exactly as the voiced sounds do in English; spellings such as ⟨p⟩ could wrongly suggest aspiration to an English speaker. [ŋ] tends to be spelled ⟨ng⟩, even though that is identical to the transcription for [ŋg]. The alveolopalatal nasal is usually written ⟨ny⟩, and writers careful to distinguish the two rhotics usually spell the trill as ⟨rr⟩. The most variation is seen in the spelling of the

alveolopalatal stop. Dixon (1972) used ⟨ɟ⟩, which at the time was the IPA symbol for [dʝ]; more common spellings include ⟨dj⟩, ⟨dy⟩, and ⟨j⟩. All of these English-like spellings, except for ⟨ɟ⟩, are common in orthographies developed for Australian languages in recent decades.

6.5.5 MORPHOLOGY

Dyirbal is a suffixing, agglutinative language. There are no prefixes. Suffixes sometimes are REDUPLICANTS, copies of the preceding morpheme. Reduplication can have different meanings in different contexts, but it typically shows some kind of intensification or augmentation of the meaning of the copied morpheme. With nouns and often with adjectives it typically expresses plural referents, which can be thought of as a sort of intensification.

- (76) a. marki~marki
 thin~AUG
 ‘very thin’
- b. ɲalŋka~ɲalŋka
 child~PL
 ‘children’

Various phonological changes may occur where a suffix is attached to a base. One of the more unusual processes is that nasals are often inserted, especially before stops. These inserted nasals agree with the following stop in place of articulation.

6.5.5.1 Nouns

GENDER

Nouns have one of four genders, or noun classes – the terms are synonymous. The genders are masculine (M), feminine (F), vegetable (VEG), and neuter (N). Gender is based on the meaning of a noun, as discussed in §6.5.6, and cannot be inferred from its pronunciation.

CASE INFLECTIONS

Dyirbal cases are marked morphologically by suffixes on the nouns, as listed in Table 6.9. When cases have multiple suffixes, the choice between them is based on the phonology of the word stem. For the genitive, the general ending is [-ŋu], but [-u] appears on stems that end in a nasal. The ergative and locative have exactly the same rules of allomorphy and result in forms that are identical except for the final vowel.

- ✕ [-ŋku] after a disyllabic stem ending in a vowel: thus ergative [jaɳaŋku] corresponds to absolutive [jaɳa] ‘man’.
- ✕ [-ku] after a trisyllabic or longer stem ending in a vowel: thus ergative [jamaniku] corresponds to absolutive [jamani] ‘rainbow’.

TABLE 6.9

Dyirbal case inflections

Case	Endings
ABS	absolutive ∅
ERG	ergative [-ŋku], [-ku], [-pu], [-tu], [-cu], [-ɽu]
DAT	dative [-ku]
GEN	genitive [-ŋu], [-u] anterior genitive [-mi]
LOC	locative [-ŋka], [-ka], [-pa], [-ta], [-ca], [-ɽa]
ABL	ablative [-ŋunu]

- ✕ a homorganic stop plus [-u] after a stem ending in a nasal or [j]. For example, the homorganic stop corresponding to the labial [m] is [p], so the ergative of [watam] ‘snake’ is [watampu]. The stop nearest to palatal [j] is [c], so the ergative of [walkuj] ‘brown snake’ is [walkujcu].
- ✕ [-ɽu] replacing a word-final liquid: thus ergative [cukumpiɽu] corresponds to absolutive [cukumpil].

As is common across languages, the absolutive case has no ending at all. Thus the stem [cukumpil-] ‘woman’ is [cukumpil] when in the absolutive case.

The ANTERIOR GENITIVE CASE genitive is used to mark a person who is not currently in possession of an object. The other cases have been discussed in connection with other languages; their special application to Dyirbal will be discussed more fully in the syntax section (§6.5.6). As we have come to expect, some morphological cases correspond to more than one syntactic function. Just as in Classical Tibetan, the ergative case doubles as an instrumental case, and the dative case can also serve an allative function, expressing the goal of motion.

NUMBER

The dual is formed by the suffix [-caran].

In a few nouns referring to people, the plural is formed by adding a suffix that varies for the different words. These number suffixes precede any case suffixes.

- (77) a. jaɽa-caran
man-DU[abs]
‘two men’
- b. jaɽa-rci
man-PL[abs]
‘more than two men’

For other nouns, as well as the adjectives that modify them, the plural is expressed by reduplicating the root of the word. When the root is reduplicated, suffixes, including case endings, are not copied.

- (78) ɲalŋka~ɲalŋka-ku
 child~PL-ERG
 ‘more than two children’

OTHER NOUN SUFFIXES

Dyirbal has many other suffixes that can be added to nouns. These usually appear before the case endings, but some can be added after them. Here is a small selection of suffixes that can be added to nouns; some of them can be added to other word classes as well.

[-ru] after a location expresses that motion is involved:

- (79) mica-ŋka-ru
 camp-LOC-motion
 ‘(going) along or through the camp’

[-ku] means ‘only’, and is always added after the case ending:

- (80) jaɬa-ŋku-pu
 man-ERG-only
 ‘only men’ (not women)

[-kapun] means ‘another’:

- (81) jaɬa-kapun
 man-another
 ‘another man’

[-ŋaru] turns a noun *N* into an adjective meaning ‘like a *N*’:

- (82) kukula-ŋaru
 platypus-like
 ‘like a platypus’

[-pil] is a **VERBALIZER**: it can turn a noun into a verb. For example, [ɲumpul] ‘beard’ becomes [ɲumpul-pil] ‘grow a beard’. A transitive verbalizer is [-mal], which has the allomorphs [-pal] or [-mpal] when added after three or more syllables. Thus [waɬu] ‘a bend’ can be turned into the verb [waɬumal] ‘make bendy’.

6.5.5.2 Determiners

Determiners (DET) in Dyirbal indicate the general location of an object referenced by a noun: its proximity, and whether it is visible. The main determiners are listed here with the glossing abbreviations we will be using for them:

- (83) jala- DET.PROX visible and close to the speaker
 ŋala- DET.NVIS not visible and far
 pala- DET visible and not close
 kila- DET.INDF indefinite location
 wuŋca- DET.Q interrogated location

Determiners agree with their governing noun in case. For the absolutive case, there is no ending to add, so [ŋala] and [pala] can appear uninflected:

- (84) pala cina
 DET[abs] foot[abs]
 ‘foot’ or ‘a foot’ or ‘the foot’

Dyirbal determiners do not indicate whether the reference is definite or indefinite – the sort of contrast afforded by the articles ‘a’ and ‘the’ in English. Therefore the articles can be added to the English translation according to the best judgment of the translator, as we showed for Russian. [jala-] has the special suppletive stem [kiŋa-] in the absolutive case, where it typically has demonstrative force: [kiŋa cina] could mean ‘this foot’ or ‘this is a foot’.

Determiners must agree with their governing noun not only in case but also in gender. Gender is indicated by adding a consonant to the determiner, after the case inflection, if there is one:

- (85) -l masculine
 -n feminine
 -m vegetable
 (nothing) neuter

The example with [cina] ‘foot’ (84) works because it is a neuter noun, and so the determiners it governs take no gender endings. Here are some examples with other genders:

- (86) a. pala-n cukumpil
 DET[abs]-F woman(F)[abs]
 b. ja-ŋku-l jaŋa-ŋku
 DET.PROX-ERG-M man(M)-ERG

Recall that gender is inherent in nouns, which is why it is glossed inside parentheses; the gender is explicitly marked only on the determiners the nouns govern. Example (86b) shows that the [-la-] of a determiner stem (except for [kila-]) can be omitted before a consonant.

The absolutive form is irregular. [kila-] is simply [kila], with no gender agreement. Other determiners have regular forms in these cases, except for the masculine. For [wuŋc-] ‘where’ the absolutive masculine is [wuŋciŋ]. The other

determiners use the ending [-ji]. Thus [paji] instead of expected *[palal]; and so [ŋaji] and [kiji].

After the gender marker may be added morphemes that show location more explicitly. These include [pajc-] ‘down’ and [taj-] ‘up’. They can be extended by a vowel that expresses distance: [-i] ‘small’ (DIM), [-a] ‘medium’, [-u] ‘large’. Thus:

- (87) pa-ŋku-l-taj-i jaʒa-ŋku
 DET-ERG-M-UP-DIM man-ERG
 ‘a man a short distance uphill’

The suffixes [kala] ‘up’ and [kali] ‘down’ have a similar meaning, and may even be combined with the above: [paŋkultajikala]. A related directional is [kalu] ‘in front’.

Some other suffixes often found on determiners include the following:

[-ja-] can be inserted before certain other case endings. Before the locative ending [-ŋka], it can express an APPREHENSIVE (APPR) case. That is, it marks a scary reason for doing something.

- (88) kalka janum pa-ji-ja-ŋka cumpulu-ka
 PROH go DET-M-APPR-loc Jumbulu(M)-loc
 ‘Don’t go for fear of Jumbulu!’

[-maŋkan] can be added to express the plural. The case ending might come before it, after it, or both.

- (89) a. pa-ŋku-maŋkan
 DET-ERG-PL
 b. pala-maŋkan-tu
 DET-PL-ERG
 c. pa-ŋku-maŋkan-tu
 DET-ERG-PL-ERG

[-cana] is an intensifier (INT) that focuses attention on a specific noun.

6.5.5.3 Pronouns

Table 6.10 lists the personal pronouns. Dyirbal does not have pronoun forms for the third person, though determiners can be used by themselves with much the same effect as a pronoun. Pronouns do not occur in the anterior genitive, locative, allative, or ablative cases.

Dyirbal pronouns have nominative-accusative alignment, unlike the ergative-absolutive alignment of nouns. Such instances of split ergativity are very common across languages.

TABLE 6.10
Dyirbal personal pronouns

Category	NOM	ACC	DAT	GEN
1SG	ŋaca	ŋajku-na	ŋajku-nku	ŋajku
1DU	ŋalici	ŋalici-na	ŋalici-nku	ŋalici-nu
1PL	ŋanaci	ŋanaci-na	ŋanaci-nku	ŋalaci-nu
2SG	ŋinta	ŋinu-na	ŋinu-nku	ŋinu
2DU	ŋupalaci	ŋupalaci-na	ŋupalaci-nku	ŋupalaci-nu
2DU	ŋuraci	ŋuraci-na	ŋuraci-nku	ŋuraci-nu

6.5.5.4 Adjectives

Adjectives take the same case endings, with the same allomorphies, as nouns.

6.5.5.5 Verbs

There are two verb paradigms in Dyirbal. Verbs must have one of the inflectional endings listed in Table 6.11. In one of the paradigms, the verb stem ends in [-l], and in the other, the verb stem ends in [-j]. The former is found more often in transitive verbs and the latter more often in intransitive verbs, but there are many exceptions to this rule, and so we gloss both endings as simply *v* for *verb*. These verbal markers are dropped in many of the inflectional categories; you can assume they are dropped unless the table lists the marker (as for the participles). These deletions are probably due in many cases to phonological simplifications. Not included in this table is the fact that even when [j] is retained for most verbs, it is often deleted when it ends up between an [i] and a consonant.

Imperatives are used for positive commands, such as [pani] ‘come!’, which is formed by dropping the verbal marker [-j] from the end of [panij]. Very short imperative forms are common among the languages of the world, but unusual among Australian languages, most of which add a suffix rather than deleting one. Negative commands use the prohibitive mood.

TABLE 6.11
Dyirbal verb inflections

Inflection	[-l] paradigm	[-j] paradigm
imperative (IMP)	∅	∅
prohibitive (PROH)	-m	-m
future (FUT)	-ŋ	-ŋ
nonfuture (NFUT)	-n	-ŋu
sequential (SEQ)	-ŋura	-ŋura
purposive (PURP)	-l-i	-(j)-ku ^a
apprehensive (APPR)	-l-pila	-mpila
relative (REL)	-ŋu-	-ŋu-
participle (PTCP)	-l-muŋa-	-j-muŋa-

^aThe [j] is deleted after [i].

TABLE 6.12
Dyirbal verb suffixes

Voice	[-l] paradigm	[-j] paradigm
Antipassive (ANTIP)	-l-(ŋ)a-j	-na-j
Instrumentive (INS)	l-ma-l	(j)-ma-l
Reflexive (REFL)	-(ji)ri-j	-j-mari-j, -j-(m)pari-j
Iterative aspect (ITER)	-kani-j	-nkani-j

Note: Elements in parentheses are sometimes omitted.

The basic indicative forms are the future and the NONFUTURE tenses – the latter being used for past or present events. The sequential and PURPOSIVE moods are used to mark clauses whose topic played a role in the preceding clause; these forms are discussed in the syntax section (§6.5.6.3). The apprehensive warns of an unpleasant result; thus it may often be translated with ‘lest’.

The endings in the table that have a hyphen after them have adjectival functions and, like adjectives, are followed by case inflections. The relative inflection is used on verbs inside relative clauses; the verb is inflected to agree with the modified noun. The morphology of the relative inflection is unusual in that the ergative and locative case endings that follow the [-ŋu-] inflection are [-ru] and [-ra], respectively, instead of the forms in [-ku] and [-ka] that appear on true adjectives. Participles describe their modified noun as being something that habitually is characterized by the verb. For example, a [cukumpil canaj-muŋa] is a woman [cukumpil] who stands around [canaj] a lot.

Verbs can also take a variety of suffixes that come before the inflectional ending. A few of these are listed in Table 6.12. Verb suffixes come after the paradigm suffix ([j] or [l]), though these are often deleted. Note that these suffixes all end in one of the two paradigm suffixes themselves, so that the suffixed verb takes inflectional endings that are appropriate for the paradigm of the suffix, not the paradigm of the basic verb. (90) shows how typical verbs ([l]-paradigm [palkal] ‘hit’ and [j]-paradigm [curkaj] ‘spear’) form the ANTIPASSIVE stem and how the nonfuture tense is formed from both stems:

- (90) a. ACT: palkal → palkan
 ↓
 ANTIP: palkalŋaj → palkalŋaju
- b. ACT: curkaj → curkaju
 ↓
 ANTIP: curkanaj → curkanaju

The antipassive, INSTRUMENTIVE, and reflexive express different voices of the verb and are discussed in §6.5.6.2. The iterative is an example of an aspect marker. Adding it to a verb adds the suggestion that the action repeats or continues.

6.5.5.6 Locational verb adjuncts

The same roots that form determiners can also take inflections that let them modify a verb, telling the location where an event took place:

(91)	Locative (LOC)	-j	place at which
	Ablative (ABL)	-ŋum	place from which
	Allative of place (ALL)	-ru, -u	place to which
	Allative of direction (ALL)	-ri, -i	direction toward which

For example, [paŋum] means ‘thence’ and [wupcaj] means ‘where’. These forms do not have any gender agreement because they modify the verb, not a noun. When these roots modify verbs, we gloss them as “place” rather than as determiners.

6.5.6 SYNTAX AND SEMANTICS

6.5.6.1 Nouns and associated elements

The gender of a noun is not tied to its phonology but to its meaning. But at first glance, the set of semantic concepts contained in each of the four noun classes appears arbitrary and haphazard:

- (92)
- a. men, kangaroos, possums, bats, most snakes, most fishes, some birds, most insects, moon, storms, rainbows, boomerangs, some spears
 - b. women, bandicoots (a kind of marsupial), dogs, platypuses, echidnas, some snakes, some fishes, most birds, fireflies, scorpions, crickets, Hairy Mary caterpillars (from *Anthela* moths), anything connected with fire or water, sun and stars, shields, some spears, some trees
 - c. vegetables, fruit, fruit trees
 - d. parts of the body, meat, bees and honey, wind, yamsticks, some spears, most trees, grass, mud, stones, noises, and language

However, Dixon (1972) showed that the concepts in these classes can all be subsumed by a few central concepts:

- (93)
- a. masculine: males; animals
 - b. feminine: females; water; fire; fighting
 - c. VEGETABLE: edible plant food
 - d. neuter: everything else

In a great number of cases, one can predict what gender a noun will have by finding the best fit to one of these categories. The word [jaŋa] means ‘man’, so it will have the masculine gender; [cukumpil] means ‘woman’ and so that word will be feminine; [pupcan] means ‘yellow walnut’, an Australian tree with edible seeds, so that word will have the vegetable gender; and [ŋacan] ‘language’ doesn’t match any of these categories, so it is neuter.

The major complication is that nouns are sometimes classified according to culturally important properties that aren't always obvious to outsiders. For example, the masculine gender is generally used for names of animals, so you might expect names of birds to be masculine. But because birds are believed to be the spirits of dead women, nouns naming most birds are feminine. Harmful things tend to get placed in that class, too, because of associations with fighting or the pain of fire. Sometimes names are assigned a gender that would seem more appropriate to something the referenced items are associated with, rather than to the items themselves. For example, spears are generally named by neuter nouns, but if the spear is of a type used for fighting, it is feminine, because fighting itself gets the feminine gender; and if the spear is the type used for spearing fish, it is masculine, because fish are masculine. And some nouns are placed in this class for unknown reasons. Across languages, this sort of categorization is typical of gender systems: each gender embodies a central key concept as well as several clusters of minor concepts that accrete through metaphor or more mysterious historical processes. The linguist and cognitive scientist George Lakoff (1987) discussed the categorization principles inherent in the Dyirbal gender system, taking his title from a synopsis of the feminine gender: *Women, fire, and dangerous things: What categories reveal about the mind*.

For the nouns that express plural number by reduplication, it is not mandatory to express number. That is, the basic unreduplicated form can be used when there are multiple referents: [palan cukumpil] could mean 'women' as well as 'a woman'.

Determiners are used with nouns that are in the absolutive, ergative, dative, or genitive cases, but do not appear with nouns that are in other cases. Nor are they used with pronouns. They generally appear before the noun. Adjectives generally come after the nouns, and they can be used with pronouns as well. As in almost all statements about Dyirbal word order, it must be remembered that speakers arrange words quite freely. Dyirbal also tolerates a good deal of ellipticality, so that either the determiner or the noun could be omitted. Thus the noun phrase [paji jaḡa pulkan] (94) could be rearranged in any order ([paji pulkan jaḡa], etc.) or stripped down to [paji pulkan] or just [pulkan]. When nouns are omitted, they can be supplied from context, or partly inferred from the gender marker on the determiner.

- (94) pa-ji jaḡa pulkan
 DET-M.ABS man(M)[abs] big[abs]
 'a big man'

Determiners and nouns must agree in case with their head noun. There is strict gender agreement between determiners and nouns as well. Number agreement between noun and adjective is optional.

Dyirbal allows verbless sentences that consist entirely of these kinds of elements. The noun phrase in (94) could stand alone as an entire sentence meaning 'The man is big'.

returned, not ‘Mother’. Thus the two clauses are linked by sharing the same topic, forming a small topic chain.

- (97) *ɲuma japu-ɲku puɟa-n | panaka-ɲu*
 father[abs] mother-ERG see-NFUT return-NFUT
 ‘Mother saw Father; he returned.’

But what if the speaker really wants to establish a subject of a transitive verb as the topic? That would be in the ergative case and therefore, by the basic rules, ineligible to be considered a topic. The rule that topics must be introduced in the absolutive can be overridden by putting the second verb in the sequential mood. That says that the preceding ergative noun is the topic and therefore should be understood to be the unexpressed absolutive argument of the second verb:

- (98) *pala juku pa-ɲku-l jaɟa-ɲku mata-n | wajɲci-ɲura*
 DET[ABS.N] stick(N)[abs] DET-ERG-M man(M)-ERG throw-NFUT go_uphill-SEQ
 ‘A man threw the stick then he went uphill.’

And what if the speaker really wants to omit the subject of a transitive verb, implying that the current topic is to be understood as the subject? For example, what if she wanted to say ‘The man brought a woman, and (she) sliced beans’, omitting the subject of the second clause? That cannot be done, because the implied subject of a transitive verb like ‘slice’ would be ergative, not absolutive. But the antipassive construction affords a way of getting around that restriction. It effectively turns the verb into an intransitive verb so that its subject – the slicer – must now be absolutive. This is done by assigning the patient – the beans – to another case not usually used for objects, often the dative (99a). The antipassive clause can now be used as a second clause in a topic chain, and the topic may be understood to be its missing subject (99b).

- (99) a. *pala-n cukumpil pa-ku-m miran-ku papi-l-ɲa-ɲu*
 DET[abs]-F woman(F)[abs] DET-DAT-VEG bean(VEG)-DAT slice-V-ANTIP-NFUT
 ‘The woman sliced beans.’
- b. *pala-n cukumpil pa-ɲku-l jaɟa-ɲku munta-n*
 DET[abs]-F woman(F)[abs] DET-ERG-M man(M)-ERG take-NFUT
- pa-ku-m miran-ku papi-l-ɲa-ɲu*
 DET-DAT-VEG bean(VEG)-DAT slice-V-ANTIP-NFUT
- ‘A man brought the woman, and she sliced the beans.’

The clear implication is that the man brought the woman somewhere so that she could scrape the beans there. The idea that a clause in a topic chain is the intended consequence of the preceding action can be made explicit by putting its verb in the *PURPOSIVE MOOD*, adding one of the purposive suffixes (Table 6.11): in this case, yielding [papi-l-ɲa-j-ku] ‘in order to slice.’

The reflexive voice works in much the same way as the antipassive, but takes a different suffix (Table 6.12):

- (100) pa-ji jaʒa caŋka-j-mari-ŋu pa-ku-m wucu-ku
 DET-ABS.M man(M)[abs] eat-V-REFL-NFUT DET-DAT-VEG fruit(VEG)-DAT
 ‘The man ate the fruit.’

One final voice to consider is the INSTRUMENTIVE. This is partly like the antipassive and the reflexive in that the normal object of an action gets knocked out of the object slot, so that it no longer takes the absolutive but the dative case. In addition, an instrument, which would normally be expressed by the ergative case (which doubles as an instrumental case in Dyrbal), is now treated as the object of the verb and therefore put into the absolutive. Therefore a sentence like ‘The man cut the tree with an ax’ would, in the instrumentive voice, put ‘man’ in the ergative – still the subject of a transitive verb – and put ‘ax’ in the absolutive – now the object of the verb. The ‘tree’ would be dative. In a topic chain, if the instrumentive verb form of ‘cut’ is used without an object, then the current topic will be understood as the instrument of cutting.

- (101) pala pari pa-ŋku-l jaʒa-ŋku maŋka-n kunpa-l-ma-l-i
 DET[N.ABS] ax(N)[abs] DET-ERG-M man(M)-ERG pick_up-NFUT cut-V-INS-V-PURP

 pa-ku juku-ku
 DET-DAT[n] tree(N)-DAT
 ‘The man picked up an ax to cut a tree (with it).’

6.5.6.4 Relative clauses

A relative clause is formed by inflecting the verb of the subordinate clause with a relativizing suffix [-ŋu-] (§6.5.5.5). To show which noun of the main clause it modifies, the relativizing suffix is followed by a case ending that agrees with that noun. Within the relative clause, that noun will be understood as the absolutive argument of the relative verb. Thus in (102a), the relative clause is [ŋinaŋu], which modifies [cukumpil] because both have the same (implicit) absolutive case, and [cukumpil] acts as an absolutive argument to the relative verb [ŋinaŋu]. Because [ŋinaŋu] is intransitive, [cukumpil] acts as its subject. In (102b), the relative clause is [wajnciŋuru], which modifies [cukumpiɰu] because both have the same case – ergative – and [cukumpiɰu] acts as an absolutive argument, that is, a subject, to the intransitive relative verb [wajnciŋuru].

If the modified noun is meant to play an ergative role within the relative clause, that is not possible per se, but an alternative voice such as the antipassive is available, just as it is used to solve a similar problem in topic chains. Thus in (102c) the verb ‘spear’ is put in the antipassive so that ‘man’ can be taken to be its subject.

- (102) a. η aca pala-n cukumpil η ina- η u pu η a-n
 1SG(NOM) DET[abs]-F woman(F)[abs] sit-REL[abs] watch-NFUT
 ‘I am watching a woman who is sitting.’
- b. pa-ji ja η a pa- η ku-n cukumpi- η u waj η ci- η u-ru pu η a-n
 DET-M.ABS man(M)[abs] DET-ERG-F woman(F)-ERG go_uphill-REL-ERG see-NFUT
 ‘The woman who was going uphill saw a man.’
- c. pa-ji ja η a paka-l- η a- η u pa-ku-l
 DET-M.ABS man(M)[abs] spear-V-ANTIP-REL[abs] DET-DAT-M
 ju η i-ku panaka- η u
 kangaroo(M)-DAT return-NFUT
 ‘The man who speared the kangaroo is coming back.’

6.5.6.5 Negation

Only verbs can be negated. The general negative particle is [kulu], whereas prohibitives are accompanied by [kalka].

- (103) η inta kalka wurpa-m
 2SG[nom] PROH speak-PROH
 ‘Don’t you speak!’

For both the imperative and the prohibitive, the second-person pronoun is optional.

Verbs may be accompanied by a noun in allative, ablative, or locative case or by a verb marker. If both are supplied, the verb marker and the locative marker have to agree in case.

6.5.6.6 Interrogative sentences

Content questions in Dyirbal contain question words: pronominal interrogatives (104a), verbal interrogatives (104b), and interrogative determiners (104c), among others.

- (104) a. waj η -cu η inu-na palka-n
 who-ERG 2SG-ACC hit-NFUT
 ‘Who hit you?’
- b. pa-ji ja η a wijama- η u
 DET-ABS.M man(M)[abs] do_what-NFUT
 ‘What is the man doing?’
- c. wu η c-in ja η a mijanta- η u
 DET.Q-ABS.M man(M)[abs] laugh-NFUT
 ‘Where is the man who is laughing?’

Polar questions may be signaled by rising intonation, as in English. Questions can also be marked by adding the enclitic [=ma] to the end of the first word of the sentence.

- (105) η inta=ma pala-n cukumpil palka-n
 2SG.NOM=Q DET-F[abs] woman(F)[abs] hit-NFUT
 ‘Did you hit the woman?’

6.5.6.7 Word order

Although, as was already pointed out, word order is quite free in Dyirbal, there are a few rules involving word order. Certain particles, such as the negatives [kulu] and [kalka], must occur somewhere before the verb they depend on, and the interrogative enclitic [=ma] must occur on the first word of the sentence. But most principles of word order are simply tendencies. (106) sketches how the dependents of a verb tend to be arranged:

- (106) time NOM-pronoun ACC-pronoun ABS-noun ERG-noun adverb verb DAT place

The basic order of the dependents of the nouns has been discussed in 6.5.6.

An implication of this ordering is that Dyirbal word order is most commonly OSV, when both arguments of the verb are nouns:

- (107) pala juku pa- η u-l ja η a- η u mata-n
 DET[ABS.N] stick(N)[abs] DET-ERG-M man(M)-ERG throw-NFUT
 ‘The man threw the stick.’

But if the subject is a pronoun, SOV order is more common:

- (108) η aca pala juku mata-n
 1SG[nom] DET[ABS.N] stick(N)[abs] throw-NFUT
 ‘I threw the stick.’

6.5.7 SAMPLE TEXT

Unlike many of the other sample texts in this book, this is a piece of unedited oral literature, recorded by a linguistic field worker (Dixon 1972: 382–387). It is an autobiographical story, in which the storyteller recounts one of her experiences as a little girl. While sitting in the camp with the tribe one night, she heard a strange noise and was told by the old people that the noise was made by the Dambun spirit.

- (109) tampun-ta η an-a- η ka mukuru jina
 Dambun(F)-LOC DET.NVIS-F.APPR-LOC quiet[abs] sit[imp]
 ‘“Sit quiet in fear of the unseen Dambun!”

The unattributed dialog in this story represents speech by the camp elders.

[η ana η ka] comes from [η an- η a- η ka]; when morphemes are combined, the second of two nasals is sometimes dropped.

- (110) wu η can-a- η ka?
 DET.Q-F.APPR-LOC
 ‘“Where is she who we are in fear of?”

- (111) ηa -n pa-n-kalu- η junu jalkaj-ca-ru wanti-n
 DET.NVIS-F[abs] DET-F-front-ABL road(N)-LOC-motion go_upriver-NFUT
 ‘‘Unseen, she is coming from out there, upriver, along the road.’’
- (112) η ira-l-pila
 pierce-V-APPR
 ‘‘She might pierce us!’’
- (113) mujmpa puni pu η a-l-pila tampun-tu
 extinguish[imp] fire(F)[abs] see-V-APPR Dambun(F)-ERG
 ‘‘Put out the fire lest Dambun see it!’’
- (114) η anaci-na cinti η ira- η
 1PL-ACC chest(N)[abs] pierce-FUT
 ‘‘She will pierce our chests!’’
 Inalienable possession, such as that involving one’s own body parts, is indicated by simply placing the words next to each other in the same case. Note that case agreement between pronouns and nouns requires negotiating two different morphosyntactic alignment systems. The absolutive is the noun case that corresponds to the pronominal accusative case.
- (115) wupca-n
 DET.Q-F[abs]
 ‘‘Where is she?’’
- (116) kila- η unta η ala-n η ala-n η anta- η mulku
 place-INDF DET.NVIS[abs]-F DET.NVIS[abs]-F call_out-FUT noise[abs]
 ‘‘Somewhere out there, unseen, she will make a noise.’’
- (117) η aca walma~walma-kali- η u pa- η um ma η ka-l-mpari-ku tucu-ku
 1SG[nom] get_up~INT-quick-NFUT place-IMP pick_up-V-REFL-PURP torch-DAT
 kinta-l-ma-l-i
 search_by_light-V-INS-V-PURP
 ‘I sprang up from there quickly to pick up a torch, so that I could use it to go searching with a light.’
 The speaker used the English word for ‘torch’, adapting it to Dyirbal phonology. It was not an established loanword.
- (118) ca η ca η aca janu
 now 1SG[nom] go[nfut]
 ‘Then I went out.’
 [janu] (as opposed to the expected *[janu-n]) is the only irregular verb in Dyirbal.
- (119) pa- η ku-ma η kan-tu kintimpa-n
 DET-ERG-PL-ERG warn-NFUT
 ‘They warned me.’

- (120) ε kalka pa-n tampun kinta-m
 INTERJECTION PROH DET-F[abs] Dambun(F)[abs] search_by_light-PROH
 pani-mpila
 come-APPR
 ‘“Hey! Don’t look for Dambun with the light, lest she come here!”’
- (121) ηaca pala kulu ηampa-n pa-ηu-manʒkan-u kuwal
 1SG[nom] DET[ABS.N] NEG listen-NFUT DET-GEN-PL-GEN speech(N)[abs]
 micu-kani-ηu
 ignore-ITER-NFUT
 ‘I didn’t listen to what they were saying; I kept ignoring them.’
- (122) pa-li-cilu ηaca janu
 place-DEM-INT 1SG[nom] go[nfut]
 pa-ku-n kinta-l-ηa-ηu tampun-ku
 DET-DAT-F search_by_light-V-ANTIP-NFUT Dambun(F)-DAT
 ‘I went in that direction, looking for Dambun with a light.’
- (123) ηaca pa-n puʒa-l-i tampun
 1SG[nom] DET-F[abs] see-V-PURP Dambun(F)[abs]
 mija-ηaru-mpa-n tampun
 what-like-V-NFUT Dambun(F)[abs]
 ‘I wanted to see Dambun, what Dambun looked like.’
- (124) pa-n-cana tampun kila-kija-n jaʒa-ηaru-mpa-n
 DET-F[abs]-INT Dambun(F)[abs] somewhere-DEM-F[abs] man-like-V-NFUT
 pala-n muraj cuta pikun-kalu~kalu
 DET-F[abs] head_hair(N)[abs] bushy[abs] nail-in_front~AUG[abs]
 marki~marki wirmpan taji~taji-kala
 thin~AUG[abs] skin_and_bones[abs] up~AUG-up[abs]
 ‘That Dambun somewhere out there [is supposed to] look something like a man,
 bushy hair, long nails, very thin, skin and bones, very very tall.’
- (125) pa-lu-pawal ηaca janu
 place-DEM-DIST 1SG[nom] go[nfut]
 ‘I went way out there.’
- (126) Jimpa pa-n-kalu puʒa-n ηaca kujipara jampi-ηu
 no DET-F-in_front see-NFUT 1SG[nom] curlew(F)[abs] fly-REL[abs]
 ‘No, I saw a curlew flying.’
 The bush stone curlew is a large nocturnal bird whose call sounds much like a wail.

- (127) kijukijukiju ηa-n-kalu pa-n jampi-n
 (cry_of_curlew) DET.NVIS-F[abs]-in_front DET-F[abs] fly-NFUT
 ‘“Keeyoo, keeyoo, keeyoo” – it flew out of sight.’
- (128) pa-ηum pala-n juŋkukan cakun-kapun-ta ηaca ηampa-n
 place-IMP DET-F[abs] another_one[abs] night-another-LOC 1SG[nom] hear-NFUT

 pa-n-kalu mulku-pi-ηku
 DET-F[abs]-in_front noise-V_{INTR}-REL[abs]

 pala-n ηampa-n kuŋu tampun
 DET-F[abs] hear-NFUT new[abs] Dambun(F)[abs]
 ‘Then on another night I heard another one; it was making noise out there;
 I heard a new Dambun.’
- (129) janu ηaca
 go[nfut] 1SG[nom]
 ‘I went out.’
- (130) maŋka-l-mpa-n ηaca tucu kinta-l-ma-l-i
 pick_up-V-INS-NFUT 1SG[nom] torch(N)[abs] search_by_light-V-INS-V-PURP
 ‘I picked up the torch, to look with it.’
- (131) tampun-ta ja-ηku-n kinta~kinta-ma-l-pila tampun
 Dambun(F)-LOC DET.PROX-ERG-F search_by_light~AUG-INS-V-APPR Dambun(F)[abs]
 ‘“What if this girl shines a light on Dambun?!”’
- (132) ηanaci caŋca manma-j-ku ja-lu-kuŋkari
 1PL[nom] now shift_camp-V-PURP place-PROX.DEM-north
 ‘“We might have to shift camp to the north.”’
- (133) pa-li-cilu ηaca janu kinta-l-ηa-ŋu
 place-DEM-INT 1SG[nom] go[nfut] search_by_light-V-ANTIP-NFUT
 ‘I went out there looking with a light.’
- (134) puɕa-n kiŋa-n-pajc-i kuku ŋina-ηu mulku-mpa-ηu
 see-NFUT DEM-F[abs]-DOWN-DIM mopoke(F)[abs] sit-REL[abs] noise-V-REL[abs]
 ‘I saw this mopoke sitting just a little bit below me, making its noise.’
 The mopoke, or southern boobook, is a small nocturnal owl.
- (135) m̄.m̄.m̄ jalama-ηu pa-n-kal-i
 (cry_of_mopoke) do_like_this-REL[abs] DET-F[abs]-DOWN-DIM

 jiti-ɣa ŋina-ηu puɣpa-jiri-ηu
 grass(N)-LOC sit-REL[abs] hide-REFL-REL[abs]
 ‘going “mm-mm-mm”, sitting, hiding down in the grass.’

- (136) kajka-pu pulkan ηaca ηaupa-n
 eye(N)[abs]-only big[abs] 1SG[nom] be_startled-NFUT
 ‘just its big eyes (visible), I jumped with fright.’
- (137) panaka-j-ara-ηu ηaca puʔa-l-a-j-ku ηujma-l-a-j-ku
 return-V-again-NFUT 1SG[nom] see-V-ANTIP-V-PURP do_properly-V-ANTIP-V-PURP
 kajka-ku pa-ku-n
 eye(N)-DAT DET-DAT-F
 ‘I went back to have a proper look at its eyes.’
 The [ŋ] of the antipassive suffix [ηaj] is often dropped.
- (138) pala-n puʔa-n kuku ηina-ηu
 DET-F[abs] see-NFUT mopoke(F)[abs] sit-REL[abs]
 ‘(I) saw a mopoke sitting there.’
- (139) caηca ηaca mijanta-ηu pala-j-cilu
 now 1SG[nom] laugh-NFUT place-LOC-INT
 ‘I laughed, right there.’
- (140) pa-ηku-maηkan-tu ηajku-na ηunca-ηu tampun-tu
 DET-ERG-many_people-ERG 1SG-ACC blame-NFUT Dambun(F)-ERG
 kitimpa-ηu
 tickle-REL[abs]
 ‘They blamed it on me being tickled by Dambun.’
- (141) pa-n ηaca pa-n mijanta-ηu pa-ku-n
 DET-F[abs] 1SG(NOM) DET-F[abs] laugh-NFUT DET-DAT-F
 tampu-ηampija-ku kuku-ku
 Dambu-whatchamacallit-DAT mopoke(F)-DAT
 ‘I was laughing at that Dambu-whatchamacallit: that mopoke.’

6.6 Exercises

6.6.1 POSSESSION MARKING IN HAWAIIAN

Reread the description of the distinction between alienable and inalienable possession in the Sketch of Hawaiian (§6.4.3.1). Contrary to our expectations, the following Hawaiian nouns are always treated as being inalienable possessions: [hale] ‘house’, [waʔa] ‘canoe’, [ʔa:ina] ‘land’, words referring to items of clothing such as [ka:maʔa] ‘shoe’, and sometimes terms for adzes. Suggest a plausible explanation for this situation.

- (1) Gender I:
- [na ke-na uβi] 'her yam'
 - [na ke-na raisi] 'her rice'
 - [na ke-na ^hdalo] 'her taro'
 - [na ke-na ika] 'her fish'
 - [na ke-na vonu] 'her turtle'
 - [na ke-na ^hbulamakau] 'her beef'
 - [na ke-na βuaka] 'her pig'
 - [na ke-na toa] 'her fowl'
- (2) Gender II:
- [na me-na ti:] 'her tea'
 - [na me-na jaŋgona] 'her kava (ceremonial drink)'
 - [na me-na moli] 'her citrus'
 - [na me-na ^hbia] 'her beer'
 - [na me-na suβu] 'her soup'
 - [na me-na uʔai] 'her water'
 - [na me-na koβe] 'her coffee'
 - [na me-na ^hbu:] 'her drinking coconut'
- (3) Gender III:
- [na no-na uʔaŋga] 'her canoe'
 - [na no-na motoka:] 'her car'
 - [na no-na βale] 'her house'
 - [na no-na tu:raŋa] 'her chief'
 - [na no-na iβola] 'her book'
 - [na no-na peni] 'her pen'
 - [na no-na ^hbilo] 'her cup'
 - [na no-na kato] 'her basket'
 - [na no-na i^hbe] 'her mat'

What determines the gender of a Fijian noun? (We are very grateful to Prof. Albert J. Schütz of the Linguistics Department at the University of Hawaii for the examples used in this exercise.)

6.6.4 BUANG MORPHOPHONEMICS

Buang (BZH) is an Austronesian language spoken in the Morobe Province of Papua New Guinea. Examine the Buang data given in Table 6.13 and then set up a single underlying shape for each morpheme – roots as well as any affixes – as well as the phonological rules by which the surface forms can be derived. You will need some critically ordered rules. Remember also that in very many languages third person has no explicit marker.

TABLE 6.13

Buang data for Exercise 4

		'my' ...	'your' ...	'his' ...
1.	'father'	amaG	amam	ama
2.	'child'	nalug	nalum	nalu
3.	'face'	malag	malam	mala
4.	'knee'	luk	lup	lus
5.	'tail'	ʒuk	ʒup	ʒus
6.	'head'	jug	jum	ju
7.	'neck'	kwag	kwam	kwa
8.	'brother'	arig	arim	ari
9.	'name'	areg	arem	are
10.	'breath'	saweg	sawem	sawen
11.	'cousin'	gadeg	gadem	gade

Note: Merrifield et al. (1987), problem 70; used with permission.

TABLE 6.14

Kiwai data for Exercise 5

Kiwai	Translation
noruso	'I eat/ate one.'
norusodurudo	'We two eat one.'
norusurudo	'We two ate one.'
roruso	'You or he eat one.'
rorusodurudo	'You two or they two eat one.'
goruso	'You or he ate one.'
nidorusori	'I shall eat one.'
nidudorusori	'We two shall eat one.'
niriso	'I eat/ate many.'
ririsodurumo	'They (more than three) eat many.'
ririso	'You or he eats many.'
norusobidurumo	'We three eat one.'
norusodurumo	'We (more than three) eat one.'

Note: Capell (1969: 73–75).

6.6.5 KIWAI MORPHOLOGY

Southern Kiwai (KJD) is a Papuan language belonging to the Trans–New Guinea group of languages and is spoken in the Western Province of Papua New Guinea.

This language has a highly complicated verb morphology. As far as it is possible on the basis of the data given, make a morphological analysis of the Kiwai verb forms in Table 6.14.

The verb morphology does not distinguish second and third persons. The system of affixation is not entirely symmetrical, and there are some morphophonemic changes involving vowels.

6.7 Suggested readings

6.7.1 GENERAL

- ❑ *Pacific languages: An introduction* (Lynch 1998). Basic introduction to the Austronesian and non-Austronesian languages of the Pacific.

6.7.2 AUSTRONESIAN

- ❑ The Austronesian dispersal and the origin of languages (Bellwood 1991).
- ❑ *The Austronesian languages* (Blust 2009). Comprehensive guide to the languages of this family, including their structure, history, and speakers.
- ❑ *The Oceanic languages* (Lynch et al. 2002). A more basic guide to the Austronesian languages of the Pacific than Blust (2009).

6.7.3 PAPUAN LANGUAGES

- ❑ *The Papuan languages of New Guinea* (Foley 1986). Description of the languages and social factors in their use, with notes on genetic classification.
- ❑ *Papuan languages of Oceania* (Wurm 1982). Highly detailed comparative study of the approximately 750 languages of this group.
- ❑ *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples* (Pawley et al. 2005). Includes a linguistic section with an introduction on classification and six chapters on Papuan languages and their history.

6.7.4 AUSTRALIAN

- ❑ *Australian aboriginal languages* (Blake 1991). A basic, very readable introduction.
- ❑ Australian Institute of Aboriginal and Torres Strait Islander Studies (<http://www.aiatsis.gov.au>). Website includes the Australian Indigenous Languages Collection, which comprises materials from 200 native languages in Australia.
- ❑ *Australian languages: Classification and the comparative method* (Bowern & Koch 2004). A set of scholarly papers on Australian languages arguing for an alternative to R. M. W. Dixon's approach to classifying them.
- ❑ *Handbook of Australian languages* (Dixon & Blake 1979–2000). Detailed grammatical sketches representing the full range of Australian languages.
- ❑ *The languages of Australia* (Dixon 2011). A detailed description of aboriginal languages with a chapter on classification.

6.7.5 DYIRBAL

- ✘ *The Dyirbal language of North Queensland* (Dixon 1972). This is the grammar on which this chapter's sketch of Dyirbal is primarily based.

6.7.6 HAWAIIAN

- ✘ *The voices of Eden: A history of Hawaiian language studies* (Schütz 1994). History of the language from its first contacts with outsiders, showing how Hawaiian was revived from near-extinction.

The Americas

In North, Central, and South America there are three groups of languages: the indigenous languages of the Americas; the languages imported from the Old World, primarily Europe, which now predominate in the region; and contact languages, which arose on the soil of the American continent. Contact languages are discussed in Chapter 8. The languages imported to the Americas from Europe have, over the years, developed traits that mark them as being somewhat different from their European antecedents; however, these differences will not be discussed here. In this chapter only the native American languages will be discussed.

The original inhabitants of the Americas came over from Asia, most likely across Beringia, the land bridge that connected Asia and America at various times throughout the Pleistocene, most recently for a long period of time ending about 16,000 years ago. This claim is supported by much genetic evidence connecting indigenous Americans with peoples of Asia (e.g. Fagundes et al. 2007; Goebel et al. 2008). It is not known just how long ago the migrations from Asia began. The bulk of the archaeological and genetic evidence points to dates between 20,000 and 15,000 years ago (the SHORT CHRONOLOGY theory). However, several sites provide tantalizing clues that human habitation may go back 40,000 years or more, a LONG CHRONOLOGY much closer to the time when Asia was settled by modern *Homo sapiens*. It is also reasonable to assume that there were many waves of migrations from Asia. Some scholars believe, for example, that an Eskimo-Aleut migration from Asia may have taken place as recently as 5,000 years ago.

North American area

This first section of the chapter discusses the languages of North America (Table 7.1), following the classification presented by Mithun (1999: 326–605). This is a fairly conservative classification, in that it does not group languages into a family unless there is a broad consensus that there is solid proof for such a grouping. We

TABLE 7.1
Some languages of North America

Name	Size	Location
Eskimo-Aleut	§7.1	Russia to Greenland
Na-Dene	§7.2	Alaska to Southwest US
Haida (HDN)	3 (§7.2)	British Columbia
Algic	§7.3	North America
Muskogean	§7.4	SE US
Yuchi (YUC)	1	Oklahoma
Siouan	§7.5	northern plains
Iroquoian	§7.6	Ontario
Caddoan	§7.7	Oklahoma
Yuman	§7.8	SW US, Baja California
Seri (SEI)	3	Sonora
Pomoan		
· Kashaya (KJU)	2	California
Palaihnihan		
· Achumawi (ACV)	2	California
Washo (WAS)	2	California
Karuk (KYH)	1	California
Yokuts (YOK)	2	California
Maidu (NMU)	0	California
Wintuan		
· Patwin (PWI)	1	California
Utian		
· Miwok (SKD)	2	California
· Costanoan	0	California
Sahaptian	§7.9	NW US
Chinookan		
· Kiksht (WAC)	0	Oregon
Zuni (ZUN)	4	New Mexico
Kiowa-Tanoan	§7.11	New Mexico
Uto-Aztecan	§7.12	Mexico
Keres (KEE)	5	New Mexico
Kutenai (KUT)	3	British Columbia
Salishan	§7.13	British Columbia
Wakashan		British Columbia
· Nootka (NUK)	3	British Columbia
· Kwakiutl (KWK)	3	British Columbia

will also mention more liberal groupings as we go along, especially those that are frequently mentioned in the literature.

These languages are mapped in Figure 7.1. You may have seen other maps giving substantially different locations for some native North American languages.

Many maps show where languages were spoken a couple of centuries ago, but we show where languages have their greatest number of speakers today. Large movements of populations can happen for many reasons, but in recent centuries the most dramatic of them were due to European settlement of the Americas. Language areas contracted or disappeared due to the death of their speakers, who succumbed to violence and to epidemics of Old World diseases to which they had no acquired immunity. In many cases, Native Americans relocated far from their homelands to avoid conflict with European settlers. In the United States, relocation was especially massive in the 1800s, when laws such as the Indian Removal Act persuaded or compelled Indians to leave their homelands for lands west of the Mississippi River, often Oklahoma.

7.1 Eskimo-Aleut

Eskimo-Aleut is the only native language family that has speakers in both Asia and America. It is divided into an Aleut branch and an Eskimo branch (Table 7.2, mapped with a “1” in Figure 7.1).

The Aleut branch of the Eskimo-Aleut family is spoken by small groups of inhabitants of various islands in the Aleutian chain and on the Commander Islands. Their self-designation is *Unangan*. There are now about 150 speakers of various Aleut dialects. All of the Aleut dialects have been heavily influenced by Russian, from which most of their technical and religious vocabulary was borrowed. There is even a language, Mednyj Aleut, that is such a thorough mixture of Aleut and Russian that many linguists classify it as a mixed language, one that is no longer clearly Eskimo-Aleut or clearly Indo-European (see §8.1 for more on mixed languages).

The Eskimo branch of the Eskimo-Aleut family consists of an Inuit branch and a Yupik branch. The term *Eskimo*, which dates back to the 1500s, is an exonym thought to originate from an Algonquian language. In the 1970s, many Canadians began to perceive the term as offensive, and the term has been controversial ever since. One theory holds that the term is offensive because it means ‘people who eat uncooked meat’, but that etymology is imaginary. More likely, the problem is that whatever name dominant groups call subordinated groups is likely to come to sound offensive sooner or later. In Canada, the endonym *Inuit* replaces the exonym *Eskimo*, but in Alaska and Siberia, many Eskimos are not Inuit, and so a more inclusive term is needed. As a result, most linguists continue to use the word *Eskimo* for this group of languages, although the longer term *Yupik-Inuit* is also available.

The Inuit branch of Eskimo stretches in a dialect chain across North America from Greenland to Alaska. Despite the inevitable encroachment from the national languages (English, French, and Danish), Inuit remains a thriving language in many areas, especially in Greenland and in Canada, where it is an official and majority language in Nunavut.



Families: ¹Eskimo-Aleut. ²Na-Dene. ³Algic. ⁴Muskogean. ⁵Siouan. ⁶Iroquoian. ⁷Caddoan. ⁸Yuman. ⁹Sahaptian. ¹⁰Tsimshianic. ¹¹Kiowa-Tanoan. ¹²Uto-Aztecan. ¹³Salishan. ¹⁴Wakashan.

TABLE 7.2
Some Eskimo-Aleut languages

Name	Size	Location
Aleut (ALE)	3	Alaska
Eskimo	6	Russia to Greenland
· Inuit	5	Russia to Greenland
· · Kalaallisut (KAL)	5	Greenland
· · Inuktitut (IKE)	5	Nunavut
· · Inuinnaqtun (IKT)	3	Northwest Territories
· · Northern Alaskan Iñupiaq (ESI)	4	Alaska
· · Seward Peninsula Iñupiaq (ESK)	4	Alaska
· Yupik	5	Russia to Alaska
· · Naukan Yupik (YNK)	2	Russia
· · Central Siberian Yupik (ESS)	4	Alaska
· · Central Alaskan Yup'ik (ESU)	5	Alaska
· · Pacific Gulf Yupik (EMS)	3	Alaska

The Yupik branch of Eskimo consists of four language groups with low mutual intelligibility, spoken in Alaska and Siberia in the Bering Sea area. Central Alaskan Yup'ik, the most widely spoken of the Yupik languages, is the dominant indigenous language of Alaska. It is described in detail in §7.26.

7.2 Na-Dene

The Na-Dene [nɑdɪ'ne] family comprises Tlingit (pronounced ['kɫɪŋkɪt] in English), which is a language spoken primarily in southern Alaska, and Athabaskan (Table 7.3, mapped with a “2” in Figure 7.1). The name combines the words for ‘people’ in Tlingit and in many Athabaskan languages. It was coined by Sapir (1915) to express his hypothesis that Tlingit and Haida were related to Athabaskan. He now appears to have been right about Tlingit and wrong about Haida. The name has stuck, although some people prefer to use the term *Athabaskan-Eyak-Tlingit* to emphasize the exclusion of Haida.

The Athabaskan branch of Na-Dene contains about 42 languages, making it one of the largest clades of languages in North America. It is named after Lake Athabasca at the border of Saskatchewan and Alberta. The internal cladistic subgrouping of Athabaskan is not well understood; consequently, it is customary to define it in terms of the three areas in which it is spoken.

The northern Athabaskan area comprises the interior of Alaska and north-western Canada. Most of the variation between Athabaskan languages is found in this area, much as we observed for Austronesian variation in Taiwan. Such a situation suggests that northern Athabaskan is unlikely to be a proper clade, but rather the homeland from which the other Athabaskan languages emerged.

TABLE 7.3

Some Na-Dene languages

Name	Size	US state or Canadian province
Tlingit (TLI)	3	Alaska
Athabaskan	6	Alaska to Southwest US
· Babine-Carrier		
· · Babine (BCR)	3	British Columbia
· · Carrier (CRX)	4	British Columbia
· Chilcotin (CLC)	4	British Columbia
· Dene		
· · Chipewyan (CHP)	4	Alberta
· · Dogrib (DGR)	4	Northwest Territories
· · Slavey (XSL)	4	Northwest Territories
· Gwich'in (GWI)	3	Northwest Territories
· Hupa (HUP)	1	California
· Southern Dene	6	
· · Navajo (NAV)	6	Arizona
· · Western Apache (APW)	5	Arizona
· · Mescalero-Chiricahua (APM)	4	New Mexico
· · Jicarilla (API)	3	New Mexico

In Alaska, most of the Athabaskan languages are either nearly extinct or moribund – spoken only by adults, with little chance of continued survival in the absence of concerted preservation campaigns. In Canada, some of the Athabaskan languages are considerably more robust. Chipewyan, whose endonym *Dëne Sų́liné* is one of several language names that contain the *Dene* element for which the language family is named, has some 11,000 speakers from Alberta to Manitoba and up into the Northwest Territories. Perhaps surprisingly from the standpoint of cultural sensitivity, *Slavey* is exactly what it looks like, the English word *slave* with the final ⟨e⟩ pronounced. The speakers call themselves *Dene* ‘people’.

Most of the Athabaskan languages spoken in Oregon and California – the Pacific group – no longer have native speakers. The few languages that have any fluent speakers includes Hupa and Tolowa (TOL) in California.

The remaining Athabaskan languages are spoken in the American southwest, quite far from the hypothesized Athabaskan homeland in Alaska. This clade is called *Southern* or *Apachean* Athabaskan. It contains perhaps the most influential Native American language in the United States: Navajo, which has over 170,000 speakers in the Arizona, Utah, and New Mexico region. It may be the only indigenous language of the United States whose native speakers are actually increasing rather than rapidly diminishing in numbers. A number of newspapers and periodicals have been published in this language. Nevertheless, there is concern that the increasing numbers of children are learning Navajo more in school than at home, a situation that augurs ill for language sustainability.

Paradoxically, the relatively large number of speakers of Navajo inspired the US Marine Corps to use it as the basis of a secret oral code during World War II. The Corps was able to recruit more than 300 Americans who were bilingual in Navajo and English; at the same time, outside the United States, Navajo was not spoken at all, nor were there published grammars or dictionaries. The code they devised consisted of codewords for a few hundred common English words plus letters of the English alphabet. All the codewords were Navajo words that had some semantic connection to the plaintext. For example, minesweeper ships were metaphorically called [tʰà:ʔ], the Navajo word for ‘beaver’. The codewords for letters were all Navajo words whose English translations started with the required letters. Thus [wólátʰí:] stood for <A> because it means ‘ant’. These semantic connections made the codewords easy for Navajo–English bilinguals to memorize and, more importantly, for them to encode and decode very quickly on the battlefield. The use of this and other codes based on minority languages is referred to as CODE TALKING. Navajo code talking was inspired by the belief that the language is so difficult to learn that the code would be unbreakable. In fact, Navajo code talking amounted to English with vocabulary substitutions, which would have made it quite vulnerable to a concerted code-breaking effort. We would normally insert here a stern warning about the dangers of exoticism – thinking that unfamiliar languages like those of Native Americans must be so unimaginably different from written Old World languages that they are not subject to commonsense reasoning – but history is often surprising. By all reports, Navajo code talking was never cracked, and it served as an enormously useful vehicle for rapidly communicating battlefield tactics in the Pacific Theater for the duration of the war.

Haida, which is spoken on Queen Charlotte Island and the opposite coast of Alaska, was once considered to be distantly related to the Na-Dene languages, but the most recent comparisons, based on a larger body of evidence, disfavor this hypothesis.

As we discussed in §4.2, a more recent proposal linking Na-Dene to Ket, the sole remaining Yeniseian language of central Siberia, has attracted much interest and tentative approval by many historical linguists.

Languages belonging to the Na-Dene family possess complicated grammars that exhibit fusional and polysynthetic traits. Many of the languages are tonal. In addition, most of these languages have complicated consonant systems that include ejectives and various complex lateral sounds, not unlike many other native languages of North America.

7.3 Algic

The Algic family contains two languages in California: Yurok and the recently extinct (1962) Wiyot (wiy). It also includes the Algonquian languages, which are found far to the east, up to and including the Atlantic coast (Table 7.4, mapped

TABLE 7.4

Some Algic languages

Name	Size	US state or Canadian province
Yurok (YUR)	2	California
Algonquian	6	Rockies to Atlantic
· Eastern Algonquian	4	Atlantic coast
· · Micmac (MIC)	4	Nova Scotia
· · Malecite-Passamaquoddy (PQM)	3	New Brunswick
· Cree (CRE)	6	Canada
· Ojibwa (OJI)	5	Canada, north US
· Menominee (MEZ)	2	Wisconsin
· Potawatomi (POT)	2	Michigan
· Fox-Sauk-Kickapoo	3	Oklahoma
· · Fox (SAC)	3	Iowa
· · Kickapoo (KIC)	3	Oklahoma
· Shawnee (SJW)	3	Oklahoma
· Blackfoot (BLA)	4	Alberta
· Arapaho (ARP)	3	Wyoming
· Cheyenne (CHY)	4	Montana

with a “3” in Figure 7.1). The latter clade is named after the Algonquian language (ALQ), which we here treat as a form of Ojibwa. Both Ojibwa and Cree have vast distributions across North America, with the expected dialectal differentiation and the expected debates about where dialect and language boundaries should be drawn. Other well-known forms of Ojibwa include Chippewa (CIW) in Michigan and Ottawa (OTW) in Ontario.

English settlement of North America began on the eastern seaboard, which was mostly Algonquian territory as far south as Virginia. Consequently, many of the most familiar Native American loanwords in English come from Algonquian languages. These include culture words like *moccasin*, *papoose*, *powwow*, *squaw*, *tomahawk*, *wampum*, and *wigwam*, plus the maize-based foods *hominy* and *pone*; but not *maize* itself, which is a Spanish loan from Taino (§7.19), nor *corn*, which is a native English word for ‘grain.’ Several names for North American animals and plants were also taken from Algonquian languages, among them *moose*, *muskrat*, *opossum*, *raccoon*, *skunk*, *hickory*, *persimmon*, and *squash*. For the most part, linguists do not know exactly what the original forms of these early loans were, but often they have been able to make Proto-Algonquian reconstructions based on reflexes of those words in well-documented Algonquian languages. Our knowledge of the protolanguage goes back to the 1920s, when Bloomfield (1925) used the comparative method to reconstruct much of it. This feat was hailed as a conclusive demonstration that the method works for all languages. Previously, there had been suggestions that American languages were so different from Old World languages that sound change might not be deterministic.

These languages introduced linguists to previously unknown types of morphosyntactic alignment. Algonquian languages have a free word order and rely on grammatical morphemes to show which nouns fulfill which grammatical relations, such as the subject and object of verbs. But the system differs from the familiar nominative-accusative systems of Russian and Finnish or the ergative-absolutive system of Dyirbal. It has three components:

First, Algonquian languages have animacy-based gender. Words for sentient beings – humans and animals – have animate gender. For the most part, nonsentient things have inanimate gender, but there are many exceptions whose motivations are not always easy to discern, such as the Blackfoot animate nouns [ist:óan] ‘knife’ and [poʔta:ʔsis] ‘stove’. Verbs take different suffixes based on whether the subject is inanimate or animate; if the latter, that suffix may also specify that the verb takes an animate or inanimate object. In (1a), [-at] marks the verb as requiring an animate object. In (1b), [-o] has the same meaning; these animacy agreement markers vary depending on the verb root. The animacy of the nouns is not overtly marked on the nouns themselves, but in this instance can correctly be inferred from their meanings. (The top line is in orthography and reflects surface pronunciation; the morphemes are glossed in their underlying forms.)

(1) Blackfoot

a. Na John iisstsímááhkatsiww amo nínaay.

an:-wa d̄ʒan i:-s:tsimá:xk-at-ji:-wa amo nína:-ji
 DEM-TOP John(AN) PST-hire-TR.AN_OBJ-DIR-TOP DEM man(AN)-OBV
 ‘John hired this man.’ (Bliss 2013: 162)

b. Omáátakohkottohkoonooka nahk Rosie anni otáni.

ot-má:t-oxkot:-oxko:n-o-ok-wa an:-wa-xk rozi
 TOP_OBJ-NEG-capable-find-TR.AN_OBJ-INV-TOP DEM-TOP-NVIS Rosie(AN)
 an:-ji w-itan-ji
 DEM-OBV 3.POSS-daughter(AN)-OBV
 ‘Her daughter can’t find Rosie.’ (Bliss 2013: 251)

Second, Blackfoot has a grammatical distinction between TOPICAL (in American language studies usually called *proximate*) and OBLVIATIVE elements in the discourse. At most one animate singular noun per clause can be marked as topical (TOP, suffix [-wa]), showing that it refers to the person or object most central to the discourse. Other nouns, including all inanimate nouns, are marked as obviative (OBV, suffix [-ji]). Thus in (1a), ‘John’ is topical, as reflected by the suffix on its accompanying demonstrative, and ‘this man’ is obviative, as shown by the suffix on the noun itself. Example (1b) is analogous for ‘Rosie’ and ‘her daughter’. The [-wa] on the verbs, though glossed as topical, can be ignored for the purpose of this discussion.

Third, transitive verbs have another suffix that tells whether the subject is higher than the object on what is somewhat misleadingly called the ANIMACY

HIERARCHY, but what is really more of an agency hierarchy. The hierarchy for Blackfoot is:

(2) 1 > 2 > TOP > OBV

That is, the speaker (first person) is more suitable as a subject of a transitive verb (agent) than is the listener, and either interlocutor is more suitable than a third party. In case both the subject and object are third person, Blackfoot ranks topical nouns above obviative ones. If this ranking expectation is met in a particular clause, the verb takes a DIRECT (DIR) suffix, which is [-ji:] when both nouns are third person; otherwise it takes an INVERSE (INV) suffix, which is [-ok] in the third person. Thus in (1a), the animacy hierarchy is followed, and the topical noun ‘John’ is the subject. In (1b), the animacy hierarchy is inverted, and the obviative noun ‘her daughter’ is the subject, despite the fact that the sentence also has a topical noun.

7.4 Muskogean

The Muskogean family (Table 7.5, mapped with a “4” in Figure 7.1) is named after Muskogee, which is also known as Creek or Seminole. At the time of European contact, this language family was spoken in Alabama and surrounding areas. Along with the language isolate Yuchi and some extinct languages, it forms the core of the Southeast linguistic area. Languages in this area are noted for consistently marking *position*, or, more accurately, *POSTURE*, especially in clauses with DURATIVE ASPECT (DUR), which marks a continuing action or a state such as location. In (3a), Choctaw literally says ‘There sits my house’, using a postural verb for inanimate (INAN) objects that are not particularly tall or long. Sentence (3b) uses a postural verb as an auxiliary. We mark these stems as inherently singular because postural verbs use suppletion to express number. For example, two people standing require the verb [hi:li], and three or more people require [joh].

TABLE 7.5
Some Muskogean languages

Name	Size	US state
Western Muskogean		
· Choctaw (CHO)	5	Mississippi
· Chickasaw (CIC)	2	Oklahoma
Alabama-Koasati		
· Alabama (AKZ)	3	Texas
· Koasati (CKU)	3	Louisiana
Mikasuki (MIK)	3	Florida
Muskogee (MUS)	4	Oklahoma

(3) Choctaw

- a. jam:akō ā-tʃokwa talá·n·ja
 there 1SG.ALIEN-house sit(SG.INAN)·DUR
 'There is my house.' (Watkins 1976: 21)
- b. dʒan at tak:on apaʃ hiki·n·ja
 John NOM apple eat stand(SG)·DUR
 'John is eating an apple.' (Watkins 1976: 22)

In 1964, about 30 linguists met at Indiana University to work out a consensus classification of the American languages. They endorsed M. Haas's proposal (1951) that Muskogean is related to four extinct language isolates of the Southeast to form a larger family that she called the *Gulf* languages. They also endorsed grouping these Gulf languages together with the Algic languages, forming an even larger family called *Macro-Algonquian*. In many additional respects that 1964 consensus was rather liberal, accepting many genetic groupings that many other academic linguists consider unproven. Indeed, a subsequent conference held in Oswego, New York, in 1976, rejected many of those groupings, more than tripling the number of clades considered unrelated to each other. Our main classification in this book follows the latter scheme, but we occasionally mention the 1964 consensus because that scheme is still widely encountered in the literature and has many proponents.

7.5 Siouan

The Siouan languages are centered on the northern prairies (Table 7.6, mapped with a “5” in Figure 7.1). This family is sometimes called *Siouan-Catawban* to emphasize that some extinct languages of the eastern United States were included in it.

TABLE 7.6
Some Siouan languages

Name	Size	US state or Canadian province
Omaha-Ponca (OMA)	2	Nebraska
Winnebago (WIN)	3	Wisconsin
Dakotan		
· Dakota (DAK)	5	South Dakota
· Lakota (LKT)	4	South Dakota
· Assiniboine (ASB)	3	Saskatchewan
· Stoney (STO)	4	Alberta
Mandan (MHQ)	1	North Dakota
Missouri River Siouan	4	
· Hidatsa (HID)	3	North Dakota
· Crow (CRO)	4	Montana

Siouan languages tend to have SOV as their basic word order. Omaha-Ponca has a typologically rare proliferation of contrasts in stops: voiced, voiceless, aspirated, and ejective. Its definite articles have different endings depending on the posture of the person it determines: standing, sitting, or moving.

7.6 Iroquoian

The Iroquoian family is divided into two branches (Table 7.7, mapped with a “6” in Figure 7.1). The family originated around the Great Lakes region, but a few languages were carried as far as the American southeast. Cherokee was located in and around northern Georgia and the Carolinas until the forced removal of Indian nations in the 1830s. Now, most speakers of Cherokee live in Oklahoma.

Iroquoian languages tend to have small inventories of consonants and few if any labials. They tend to have one or two nasalized vowels. Verbs have elaborate morphological structure. In Cherokee, for example, all verbs contain a prefix that agrees with the subject and object in person and number. In (4), that is [n-], which is phonologically reduced from [i:ni:]. Verbs also must contain a suffix that marks aspect: here, [sk-], which denotes an uncompleted action. There are also possibilities for many other affixes, as well as incorporated nouns.

(4) Cherokee

SƏʒʔəʔAT.

té:-n-asú:lé:-sk-óʔi

ITER-1DU-wash_hands-IPFV-HAB

‘You and I regularly wash our hands.’ (Montgomery-Anderson 2008: 57)

TABLE 7.7

Some Iroquoian languages

Name	Size	US state or Canadian province
Cherokee (CHR)	5	Oklahoma
Northern		
· Tuscarora (TUS)	1	Ontario
· Iroquois		
· · Cayuga (CAY)	3	Ontario
· · Mohawk (MOH)	4	New York
· · Oneida (ONE)	3	Ontario
· · Onondaga (ONO)	2	Ontario
· · Seneca (SEE)	3	New York

7.7 Caddoan

Caddoan was originally represented by languages covering the central United States from North Dakota to northern Louisiana and Texas, but now only four languages survive (Table 7.8, mapped with a “7” in Figure 7.1).

The earlier consensus lumped the Siouan, Iroquoian, and Caddoan languages, as well as the Yuchi isolate, together into a large family called *Macro-Siouan*.

7.8 Yuman

The Yuman languages are spoken in the US southwest and in Baja California (Table 7.9, mapped with a “8” in Figure 7.1). This family is sometimes called *Cochimí-Yuman* to emphasize that the extinct Cochimí (COJ) of Baja California belonged to it.

Along with Seri, Pomoan, Achumawi, Washo, Karuk, some extinct languages, and a few others discussed below under Mesoamerica (Tol, Tlapanec, and

TABLE 7.8
Some Caddoan languages

Name	Size	State in US
Northern Caddoan		
· Arikara (ARI)	1	North Dakota
· Pawnee (PAW)	2	Oklahoma
· Wichita (WIC)	1	Oklahoma
Caddo (CAD)	2	Oklahoma

TABLE 7.9
Some Yuman languages

Name	Size	State in Mexico or US
Delta-Californian		
· Cocopa (COC)	3	Baja California
· Kumiai (DIH)	3	Baja California
River		
· Mojave (MOV)	3	Arizona
· Quechan (YUM)	3	California
· Maricopa (MRC)	3	Arizona
Pai		
· Upland Yuman (YUF)	4	Arizona
· Paipai (PPI)	2	Baja California

Tequistlatecan), the 1964 consensus grouped Yuman into a large theoretical family called *Hokan*.

SWITCH-REFERENCE systems were first documented in Hokan languages (Jacobsen 1967). These are systems that grammatically indicate whether the subjects of two consecutive clauses refer to the same entity (SS) or a different one (DS). These two situations are marked by two separate suffixes in Yavapai (5).

(5) Yavapai dialect of Upland Yuman

a. tokatoka-t̃j savakjuva u-t-k t̃jikwar-kiñ
Tokatoka-NOM Savakyuva see-time-SS laugh-PFV
‘When Tokatoka looked at Savakyuva, he (Tokatoka) laughed.’

b. tokatoka-t̃j savakjuva u-t-m t̃jikwar-kiñ
Tokatoka-NOM Savakyuva see-time-DS laugh-PFV
‘When Tokatoka looked at Savakyuva, he (not Tokatoka) laughed.’ (both from
Finer 1985: 37).

7.9 Sahaptian

The Sahaptian languages are spoken in the plateau at the juncture of Washington, Oregon, and Idaho (Table 7.10, preceded by “9” in Figure 7.1).

Nez Perce has been analyzed as having an unusual type of morphosyntactic alignment called TRIPARTITE ALIGNMENT (Rude 1986). As we have seen, most languages that have morphological case either treat the subject of intransitive verbs the same as the subject of transitive verbs – nominative-accusative languages – or the same as the object of transitive verbs – ergative-absolutive languages. Nez Perce leaves the subject of an intransitive verb unmarked but has special suffixes for the other two relations, so that, effectively, there are three separate forms for the three grammatical relations (6).

TABLE 7.10
Some Sahaptian languages

Name	Size	State in US
Nez Perce (NEZ)	3	Idaho
Sahaptin		
· Tenino (TQN)	2	Oregon
· Umatilla (UMA)	2	Oregon
· Walla Walla (WAA)	2	Oregon
· Yakima (YAK)	2	Washington

(6) Nez Perce

a. hi-páaj-na háama
 3.V_{INTR} -arrive-PFV man[abs]

‘The man arrived.’

b. háama-nm pée-ʔwi-je wewúkije-ne
 man-ERG 3.TR-shoot-PFV elk-ACC

‘The man shot the elk.’ (both from Rude 1986: 126)

7.10 Tsimshianic

Tsimshianic languages are spoken in British Columbia and Alaska (Table 7.11, preceded by a “a” in Figure 7.1).

The 1964 consensus, endorsing a theory proposed a century ago, grouped together into a family called *Penutian* all the language groups from Yokuts to Zuni in Table 7.1, plus a few Mesoamerican groups (Mixe-Zoque, Mayan, Totonacan, and Huave).

Yokuts, which is either a language family (Yokutsan) or a language isolate, depending on how one regards its member language varieties, has a complicated phonology. This richness has made it a traditional focus of research and a testbed for different phonological theories.

7.11 Kiowa-Tanoan

Kiowa-Tanoan includes the languages listed in Table 7.12 (preceded by a “b” in Figure 7.1).

These languages have an unusual way of marking number on nouns. In Kiowa (7), all nouns in their basic, unsuffixed form (left side of table), can be interpreted as having dual number. In addition, almost all nouns can express at least one other number in their basic form. For animates like ‘cow’, the basic form can also express singular; thus [tʰsɛnbó:] means ‘one or two cows’. Most inanimates are like ‘ribbons’, whose basic form can also express a plural (three or more in Kiowa), so that

TABLE 7.11

Some Tsimshianic languages

Name	Size	Province in Canada
Tsimshian (TST)	3	British Columbia
Nass-Gitksan		
· Nisga’a (NCG)	3	British Columbia
· Gitksan (GTT)	3	British Columbia

TABLE 7.12
Some Kiowa-Tanoan languages

Name	Size	State in US
Kiowa (KIO)	3	Oklahoma
Tiwa		
· Northern Tiwa (TWF)	4	New Mexico
· Southern Tiwa (TIX)	4	New Mexico
Tewa (TEW)	4	New Mexico
Jemez (TOW)	4	New Mexico

[ʒlp^hã:] means ‘two or more ribbons’. Some others, mostly fruit like ‘apples’, can express only a dual, and other inanimates can represent any number in their basic form. To express any number other than that expressed by a noun’s basic form, an inverse morpheme is added. Thus [t̄senbó:-gɔ] means ‘any number of cows other than two or one’, thus three or more; [ʒlp^hã:-gɔ] means ‘any number of ribbons other than two or more than two’, thus one; and so forth. The inverse morpheme is never used with words like [tó] that can express any number in their basic form.

(7) Kiowa (Wonderly et al. 1954)

[t̄senbó:] ‘cow(s)’	DU or SG	[t̄senbó:-gɔ]	PL
[ʒlp ^h ã:] ‘ribbons’	DU or PL	[ʒlp ^h ã:-gɔ]	SG
[álɔ:] ‘apples’	DU	[álɔ:-gɔ]	SG or PL
[tó] ‘house(s)’	DU or SG or PL	—	

7.12 Uto-Aztecan

Table 7.13 lists some members of the Uto-Aztecan family; in Figure 7.1, the codes for these languages are preceded by a “c”. Tepehuan languages are tonal, but most other Uto-Aztecan languages are not. The Aztecan subbranch contains a large number of languages, including Classical Nahuatl, which was the chief language of the Aztec empire at the time of Spanish conquest. The modern languages belonging to this subgroup are all descendants of Nahuatl, which served as a lingua franca throughout central Mexico. Over two dozen varieties have been recognized by various researchers, many of them being quite divergent from each other. In Table 7.13 we list both the most populous Nahuatl language – dialects of Huasteca Nahuatl are spoken by about a million people – and the most linguistically divergent, Pipil, which is spoken in El Salvador.

The grammatical structure of Uto-Aztecan languages is less elaborate than, for example, that of the Na-Dene languages. Most of the languages are agglutinative in type and have elaborate suffixation. The Numic languages are strongly verb-final, and it is believed that this feature ought to be reconstructed for Proto-Uto-Aztecan.

TABLE 7.13
Some Uto-Aztecan languages

Name	Size	State in US or Mexico
Numic		
· Western Numic		
· · Mono (MNR)	2	California
· Central Numic		
· · Shoshone (SHH)	4	Idaho
· · Comanche (COM)	3	Oklahoma
· Southern Numic		
· · Kawaiisu (XAW)	1	California
· · Ute-Chemehuevi (UTE)	3	Utah
Tübatulabal (TUB)	1	California
Takic		
· Cahuilla (CHL)	2	California
· Luiseño (LUI)	1	California
Hopi (HOP)	4	Arizona
Taracahitic		
· Cahitan		
· · Mayo (MFY)	5	Sonora
· · Yaqui (YAQ)	5	Sonora
· Tarahumaran		
· · Guarijío (VAR)	4	Chihuahua
· · Tarahumara (TAR)	5	Chihuahua
Corachol		
· Cora (CRN)	4	Nayarit
· Huichol (HCH)	5	Jalisco
Tepiman		
· O'odham (OOD)	5	Arizona
· Pima Bajo (PIA)	3	Chihuahua
· Northern Tepehuán (NTP)	4	Chihuahua
· Southern Tepehuán (STP)	5	Durango
Aztecan		
· Classical Nahuatl (NCI)	0	México
· Eastern Huasteca Nahuatl (NHE)	6	Hidalgo
· Pipil (PPL)	2	El Salvador

Whorf (1956) made Hopi a major focus of his work on linguistic relativity. He claimed, among other things, that Hopi does not have tenses for expressing past, present, and future. Such linguistic properties, he believed, had a profound effect on its native speakers, causing them to conceptualize time in a completely different way from the European notion of time as

a linear fourth dimension. Most linguists could make little sense of Whorf's statements – Hopi's treatment of time is quite typical of languages around the world (Malotki 1983) – but the public was eager to believe that the Hopi people live in an unimaginably exotic world where time as we understand it does not exist.

The 1964 consensus grouped Kiowa-Tanoan and Uto-Aztecan together into a group called *Aztec-Tanoan*.

7.13 Salishan

Most Salishan languages (Table 7.14; preceded by a “^d” in Figure 7.1) are spoken only by small numbers of speakers, most of whom are middle-aged or older.

Salishan languages have very rich consonantal systems. Shuswap, for example, uses several places of articulation, including the typologically rare pharyngeals, and a contrast between velars and uvulars. It also has glottalized or ejective versions of almost all of its consonants. Salishan languages often permit long clusters of consonants. Words such as the example in (8) have been used to test and extend theories about syllable structure.

TABLE 7.14
Some Salishan languages

Name	Size	US state or Canadian province
Bella Coola (BLC)	2	British Columbia
Central Salishan		
· Comox (COO)	2	British Columbia
· Sechelt (SEC)	2	British Columbia
· Halkomelem (HUR)	3	British Columbia
· Squamish (SQU)	1	British Columbia
· Straits Salish (STR)	2	British Columbia
· Lushootseed (LUT)	2	Washington
Interior Salishan		
· Coeur d'Alene (CRD)	1	Idaho
· Columbian (COL)	2	Washington
· Kalispel-Salish (FLA)	2	Montana
· Spokane (SPO)	1	Washington
· Okanagan (OKA)	3	British Columbia
· · Lillooet (LIL)	3	British Columbia
· · Shuswap (SHS)	3	British Columbia
· · Thompson (THP)	3	British Columbia

(8) Bella Coola

 $\text{t}\chi^{\text{wt}}\text{-}\text{t}\text{-}\widehat{\text{ts}}\text{-}\text{x}^{\text{w}}$

spit-PST-1SG-2SG

'You spat on me.' (Bagemihl 1998: 74)

The grammar of Salishan languages is complex and polysynthetic. They make extensive use of reduplication to express grammatical functions such as plural number: Shuswap [$^{\text{h}}\text{p}\text{e}\text{s}\text{ə}\text{t}\text{k}^{\text{w}}\text{e}$] 'lake', [$\text{p}\text{ə}\text{s}'\text{p}\text{e}\text{s}\text{ə}\text{t}\text{k}^{\text{w}}\text{e}$] 'lakes' (van Eijk 1998). They have many suffixes, including a large number of LEXICAL SUFFIXES: suffixes that express basic noun-like meanings such as body parts but that appear to be etymologically unrelated to noun roots of the same meaning.

Even the very liberal consensus of 1964 did not group these last-mentioned languages and families – Keres through Wakashan in Table 7.1 – into larger groups. The general consensus is that any similarities among them is due to extensive borrowing. In particular, Salishan and Wakashan (which is preceded by an “e” in Figure 7.1) exhibit a number of similarities in structure and phonology as part of a Pacific Northwest sprachbund. But there are important classifications that assigned broader genetic relations between these languages and others. In 1929 Sapir was commissioned to produce a classification for the *Encyclopædia Britannica*. He developed a scheme that grouped all of the languages of North America into six families, for example grouping Keres in Hokan and many of the remaining languages as sisters of Algonquian. Greenberg (1987) went even further, with a theory that grouped all languages of the Americas but Eskimo-Aleut and Na-Dene into one huge family called *Amerind*. His conclusions were based on subjectively judging similarities between words in different languages. Because these data were not measured or assessed by any objective standard, his claim was not subject to replication and has received little acceptance. See, for example, the very negative review of Greenberg's Amerind hypothesis by L. Campbell (1988) and Kaufman's (1990: 15–17) comments on Greenberg's methodology.

Mesoamerican area

Mesoamerica – the region from central Mexico to northern Costa Rica – is traditionally separated from the rest of North America in part because of its distinctive culture. It was, for example, one of the very few places in the world where writing was invented independently. It is also a sprachbund, where hundreds of languages exerted much influence on each other. That said, Mesoamerica was not a cultural island, and some northern language families extend into Mesoamerica, notably Uto-Aztecan (§7.12), as do some language families that will be discussed below under South America: Chibchan and Misumalpan. See Figure 7.2 for locations. The classification used for the Mesoamerican languages (Table 7.15) follows that of L. Campbell 1997.



FIGURE 7.2 Some languages of Mesoamerica. Families: ¹Oto-Manguean. ²Totonacan. ³Mixe-Zoquean. ⁴Mayan. ⁵Tequistlatecan.

TABLE 7.15

Some indigenous languages of Mesoamerica

Name	Size	Location
Oto-Manguean	§7.14	Mexico
Totonacan	§7.15	Mexico
Mixe-Zoquean	§7.16	Mexico
Mayan	§7.17	Guatemala
Tol (ITC)	3	Honduras
Tarascan		
· Western Highland Purepecha (PUA)	6	Mexico
Tequistlatecan		
· Lowland Oaxaca Chontal (CLO)	4	Mexico
· Highland Oaxaca Chontal (CHD)	4	Mexico
Huave (HUV)	5	Mexico

7.14 Oto-Manguean

The languages belonging to this family are currently spoken in Mexico, primarily in Oaxaca (Table 7.16; labeled with “1” in Figure 7.2). Oto-Manguean languages are tonal and most have nasalized vowels and VSO word order.

In the Otomi counting system, there are basic words for 1, 2, 3, 4, 5, 10, and 20, and other numbers are formed by adding or multiplying smaller units. Thus 57 is literally ‘ $(2 \times 20) + 10 + (2 + 5)$ ’ (Gilsdorf 2012: chap. 9). Thus it has a mixed based system of the pattern 5–10–20. The inclusion of 20 as a base – the VIGESIMAL component – is typical of languages throughout Mesoamerica, but is also found in a large number of languages throughout the world, including, at least in part, European languages like French, Danish, and Welsh.

7.15 Totonacan

The Totonacan family has an especially complicated morphology, resulting in quite long words. It is divided into two groups, Totonac and Tepehua. The traditional nomenclature treats these as two languages having several dialects (e.g. Coyutla Totonac, Northern Totonac, etc.), but the current approach is to treat many of these dialects as comprising a dozen or more different languages, several of which are listed in Table 7.17 (indicated by the superscript “2” in Figure 7.2).

7.16 Mixe-Zoquean

The Mixe-Zoquean languages (“3” in Figure 7.2) are spoken in the Isthmus of Tehuantepec, the strip of land where the Gulf of Mexico most closely approaches the Pacific Ocean. Table 7.18 presents the most widely accepted grouping of these

TABLE 7.16
Some Oto-Manguean languages

Name	Size	State in Mexico
Eastern Oto-Manguean		
· Amuzgo-Mixtecan		
· · Amuzgo (AMU)	5	Guerrero
· · Mixtecan		
· · · Cuicatec (CUX)	4	Oaxaca
· · · Mixtec	6	Oaxaca
· · · · Jamiltepec (MXT)	4	Oaxaca
· · · · Juxtlahuaca (VMC)	5	Oaxaca
· · · · Metlatónoc (MXV)	5	Guerrero

(continued)

TABLE 7.16

Continued

Name	Size	State in Mexico
· · · · Northern Tlaxiaco (XTN)	5	Oaxaca
· · · · San Juan Colorado (MJC)	4	Oaxaca
· · · · Yosondúa (MPM)	4	Oaxaca
· · · · Silacayoapan (MKS)	5	Oaxaca
· Popolocan-Zapotecan		
· · Popolocan		
· · · Mazatec	6	
· · · · Huautla (MAU)	5	Oaxaca
· · · · Jalapa de Díaz (MAJ)	5	Oaxaca
· · · · San Jerónimo Tecóatl (MAA)	5	Oaxaca
· · · Western Popoloca (PCA)	4	Puebla
· · Zapotecan		
· · · Chatino		
· · · · Nopala (CYA)	4	Oaxaca
· · · · Zenzontepec (CZN)	4	Oaxaca
· · · Zapotec (ZAP)		
· · · · Miahuatlán (ZAM)	4	Oaxaca
· · · · Tlacolulita (ZPK)	3	Oaxaca
· · · · Choapan (ZPC)	5	Oaxaca
· · · · Isthmus (ZAI)	5	Oaxaca
· · · · Ocotlán (ZAC)	5	Oaxaca
Western Oto-Manguean		
· Tlapanec (TPX)	5	Guerrero
· Oto-Pame-Chinantecan		
· · Chinantecan		
· · · Lalana-Tepinapa (CNL)	5	Oaxaca
· · · Ojitlán (CHJ)	5	Oaxaca
· · · Palantla (CPA)	5	Oaxaca
· · · Quiotepec (CHQ)	4	Oaxaca
· · · Usila (CUC)	4	Oaxaca
· · Oto-Pamean		
· · · Chichimeca-Jonaz (PEI)	4	Guanajuato
· · · Otomi	6	
· · · · Eastern Highland (OTM)	5	Hidalgo
· · · · Mezquital (OTE)	5	Hidalgo
· · · · Querétaro (OTQ)	5	Querétaro
· · · · Temoaya (OTT)	5	México
· · · · Tenango (OTN)	5	Hidalgo
· · · Mazahua (MAZ)	6	México
· · · Pame		
· · · · Central Pame (PBS)	4	San Luis Potosí
· · · · Northern Pame (PMQ)	4	San Luis Potosí

TABLE 7.17
Some Totonacan languages

Name	Size	State in Mexico
Totonac	6	
· Coyutla Totonac (TOC)	5	Puebla
· Filomeno Mata-Coahuilán (TLP)	5	Veracruz
· Northern Totonac (TOO)	4	Puebla
· Papantla (TOP)	5	Veracruz
· Sierra Totonac (TOS)	6	Puebla
Tepehua		
· Huehuetla (TEE)	4	Hidalgo
· Pisaflores (TPP)	4	Puebla
· Tlachichilco (TPT)	4	Veracruz

TABLE 7.18
Some Mixe-Zoquean languages

Name	Size	State in Mexico
Mixean	6	
· Oluta Popoluca (PLO)	3	Veracruz
· Sayula Popoluca (POS)	4	Veracruz
· Oaxaca Mixean		
· · Lowland Mixe (MIR)	5	Oaxaca
· · Midland Mixe		
· · · South Midland Mixe (MXQ)	4	Oaxaca
· · South Highland Mixe	5	Oaxaca
· · · Zempoaltepetl (MXP)	4	Oaxaca
· · North Highland Mixe (MTO)	4	Oaxaca
Zoquean		
· Chiapas Zoquean		
· · Central Zoque (ZOC)	5	Chiapas
· · North Zoque (ZOS)	5	Chiapas
· · Northeast Zoque (ZOR)	4	Chiapas
· Chimalapa Zoquean		
· · Santa María Chimalapa Zoque (ZOH)	4	Oaxaca
· Gulf Zoquean		
· · Sierra Popoluca (POI)	5	Veracruz

languages, that of Wichmann (1995). The traditional taxonomy treats this family as comprising three languages, Mixe, Zoque, and Popoluca, but recent approaches distinguish at least a dozen languages.

L. Campbell & Kaufman (1976) have argued that a Mixe-Zoquean language, possibly Proto-Mixe-Zoquean, was spoken by the Olmecs, the foundational

civilization of Mesoamerica, which dates back to about 1500 BC. They based their argument on comparisons of linguistic and archaeological data. Items that were introduced or first raised to importance by the Olmecs, such as plants like cacao and tomato, are often named throughout Mesoamerica by terms that are best explained as borrowings from Proto-Mixe-Zoquean. For example, 'cacao' is reconstructed as Proto-Mixe-Zoquean [kakawa], which appears as [kakawa] in Nahuatl, and in similar forms in several other unrelated languages. Clearly, borrowing is involved, but how do we know the word originally came from Proto-Mixe-Zoquean and not, say, Proto-Uto-Aztecan, the ancestor of Nahuatl? One piece of reasoning is that other Uto-Aztecan words do not have forms like [kakawa-], so it cannot be traced back to Proto-Uto-Aztecan. Forms very similar to [kakawa-] do appear throughout most of the Mayan languages, but Mayan languages have mostly monosyllabic roots, so it would be odd for the word [kakawa-] to have originated in Proto-Mayan: it sounds like a loanword. Archaeological linguistics can be very complicated and fraught with uncertainties, and so proposals such as that of Campbell and Kaufman are usually considered intriguing hypotheses rather than established fact.

7.17 Mayan

Table 7.19 is based on L. Campbell & Kaufman's (1985) cladistic analysis of the Mayan language family (plotted with a prefixed "4" in Figure 7.2). Ancient Mayas were one of only a handful of civilizations to independently invent writing, and the only native American people who did so before the coming of the Europeans. This script, which has been deciphered recently, began as a logographic system that developed a syllabic system. As in many other syllabographies, most Mayan syllabograms represented a consonant followed by a vowel (CV). Closed syllables (CVC) were written with two CV signs, in which case the vowel of the second sign was silent but was chosen to match the vowel of the root. For example, the Yucatec word [ku:ts̄] 'turkey' was spelled as if [ku:ts̄u]. The Maya script was first used by Cholan speakers, who were the principal bearers of the Classic Lowland Maya culture (300–900 AD), and later by Yucatecans, mainly to record dynastic histories.

South American area

An estimated 11.2 million people in South America speak an American native language. Unfortunately, only in the latter half of the 20th century did the study of South American languages begin to make significant progress. As a result, the genetic classification of these languages is still in quite a primitive state. At

TABLE 7.19

Some Mayan languages

Name	Size	Country
Huastec (HUS)	6	Mexico
Yucatecan–Core Mayan		
· Core Mayan		
· · Cholan-Tzeltalan		
· · · Cholan		
· · · · Chol-Chontal		
· · · · · Chol (CTU)	6	Mexico
· · · · · Chontal (CHF)	5	Mexico
· · · Tzeltalan		
· · · · Tzeltal (TZH)	6	Mexico
· · · · Tzotzil (TZO)	6	Mexico
· · Q'anjob'alan-Chujean		
· · · Chujean		
· · · · Chuj (CAC)	5	Guatemala
· · · · Tojolabal (TOJ)	5	Mexico
· · · Q'anjob'alan		
· · · · Q'anjob'al-Akateko-Jakalteko		
· · · · · Akateko (KNI)	5	Guatemala
· · · · · Jakalteko (JAC)	4	Guatemala
· · · · · Q'anjob'al (KJB)	5	Guatemala
· · · · Mocho (MHC)	3	Mexico
· · K'ichean-Mamean		
· · · Mamean		
· · · · Awakateko-Ixil		
· · · · · Awakateko (AGU)	5	Guatemala
· · · · · Ixil (IXL)	5	Guatemala
· · · · Teco-Mam		
· · · · · Mam (MAM)	6	Guatemala
· · · K'ichean		
· · · · Poqom-K'ichean		
· · · · · Core K'ichean		
· · · · · · Kaqchikel-Tz'utujil		
· · · · · · · Kaqchikel (CAK)	6	Guatemala
· · · · · · · Tz'utujil (TZJ)	5	Guatemala
· · · · · · K'iche' (QUC)	7	Guatemala
· · · · · Poqom		
· · · · · · Poqomchi' (POH)	5	Guatemala
· · · · Q'eqchi' (KEK)	6	Guatemala
· Yucatecan		
· · Yucatec-Lacandon		
· · · Yucatec (YUA)	6	Mexico
· · · Lacandón (LAC)	3	Mexico

TABLE 7.20

Indigenous languages of South America

Name	Subsection
Intermediate area	§7.18
Western Amazonia	§7.19
Northern foothills	§7.20
Andes	§7.21
Southern foothills	§7.22
South	§7.23
Central Amazonia	§7.24
Northern Amazonia	§7.25

one extreme, we have Greenberg's (1987) Amerind hypothesis that lumps all South American Indian languages into one family. At the other extreme, we have Kaufman's (1990) classification, which lists 118 families and language isolates. In between, there are many proposals for how some of Kaufman's groups might be combined into larger groups. Most authorities are agreed that many of the larger groups are supported by intriguing pieces of evidence, but the consensus is that these larger groupings are hypotheses that have not yet been proved. The following presentation of genetic relationships is based on the careful and conservative work of Kaufman, taking into account updates in sources such as L. Campbell (1997). In this overview (Table 7.20, Figure 7.3), we do not strive to be exhaustive; in particular, we have omitted most languages that are not known to have any native speakers. Campbell, Kaufman, and comprehensive sources like *Ethnologue* provide more exhaustive lists of current and historical languages.

7.18 Intermediate area

Because of the large number of language families in South America, we have arranged this survey geographically in relatively short sections. This first section is for languages centered in the northwest of South America and parts of Central America (Table 7.21). This area is called *intermediate* because it lay between the Maya civilization to its north and the Inca civilization to its south. We then tour the continent in a counterclockwise direction.

The Chocoan languages are spoken in Colombia and Panama (plotted with a prefixed "1" in Figure 7.3). There is broad support for the idea that Paezan is distantly related to the Barbacoan family (prefixed with "2" in Figure 7.3), which extends from Colombia south into Ecuador. The Chibchan family stretches from Honduras to Venezuela ("3" in Figure 7.3). It is often hypothesized to be connected to the Misumalpan family ("4"), which is spoken in Honduras and Nicaragua.



FIGURE 7.3 Some languages of South America. Families: ¹Chocoan. ²Barbacoan. ³Chibchan. ⁴Misumalpan. ⁵Arawakan. ⁶Tucanoan. ⁷Jivaroan. ⁸Cahuapanan. ⁹Zaparoan. ^aWitotoan. ^bAymaran. ^cChipaya-Uru. ^dPanoan. ^eTacanan. ^fMatacoan. ^gGuaykuruan. ^hMascoyan. ⁱZamucoan. ^jJean. ^kJabutian. ^lNambiquaran. ^mTupian. ⁿCariban. ^oSálivan. ^pQuechuan.

7.19 Western Amazonia

A selection of Western Amazonian languages is presented in Table 7.22. The Arawakan family is the largest family of indigenous languages in the Americas, both in geographical coverage – it is spoken in almost every country of South America (plotted with a prefixed “5” in Figure 7.3) – and in the number of languages it contains. It includes the extinct language Taino, which, as the dominant language in the Caribbean, was recorded by Columbus and is the source of the first American loanwords into Spanish and thence into many other languages; among

these borrowings, cited here in their Spanish form, are *canoa* ‘canoe’, *cazabe* ‘casava’, *hamaca* ‘hammock’, *huracán* ‘hurricane’, and *maíz* ‘maize’.

The languages belonging to the Tucanoan family are found in Colombia, Ecuador, Peru, and Brazil (plotted with a prefixed “6” on the map in Figure 7.3). Almost all of those languages are SOV. The American continent in general has a disproportionately high number of languages that have nasalized vowel phonemes, but the equatorial area is particularly abundant in them. Many of these languages have NASAL HARMONY. In Cubeo, if any phoneme in a syllable is phonologically nasal, then every voiced phoneme in the syllable must be nasal. That is, there are syllables like [ba], [mã], and [pã], but not *[bã] or *[ma]. Such a pattern leads many phonologists to say that in this language, nasality is a feature that belongs to the syllable rather than to the phoneme. Thus it is an important example of a PROSODIC feature, one that is best understood in terms of linguistic structures that are larger than segments. If any morpheme ends in a nasalized syllable, the nasality spreads to all suffixes that follow that syllable. Thus in (9a), both words have the same underlying suffixes (second line), but their surface forms (top line) are different, because the nasal phoneme has spread from the root ‘man’. The spread is blocked by voiceless consonants, which cannot be nasalized: [k] in (9b).

- (9) Cubeo (Chacon 2012: Chap. 3)
- a. [dʒa'vi-βa-ɾe] [ĩmĩ-wã-ɾẽ]
 /jawi-wa-de/ /'ibi-wa-de/
 jaguar-AN.PL-OBL man-AN.PL-OBL
 ‘for the jaguars’ ‘for the men’
- b. [a-'biko] [ã-'miko]
 /a-biko/ /ã-biko/
 eat-3.F say-3.F
 ‘She ate.’ ‘She said.’

7.20 Northern foothills

Table 7.23 lists some of the languages of the northern foothills of the Andes.

The Jivaroan family is spoken in Ecuador and Peru (prefixed with a “7” in Figure 7.3). The Cahuapanan family of Peru (prefixed with an “8” on that map) consists of two mutually unintelligible languages. The Zaparoan languages (prefixed “9”) are found in Ecuador and Peru. The Witotoan family (prefixed “a”) includes a few languages of Colombia, Peru, and Brazil.

Yagua is the sole surviving member of the Peba-Yaguan family, which had two other members in historic times. It has about 40 classifiers that must be used with numbers. Like many other South American languages, it is mildly polysynthetic; verbs in particular can take many suffixes, though the great bulk of verbs rarely

TABLE 7.21

Some languages of the Intermediate Area

Name	Size	Country
Chocoan		
· Emberá-Catío (CTO)	5	Colombia
· Northern Emberá (EMP)	5	Colombia
· Southern Emberá (CMI)	5	Colombia
Paezan		
· Páez (PBB)	5	Colombia
Barbacoan		
· Northern Barbacoan		
· · Awa-Cuaiquer (KW1)	5	Colombia
· Southern Barbacoan		
· · Chachi (CBI)	4	Ecuador
Chibchan (CBA)		
· Chibchan A		
· · Ngäbere (GYM)	6	Panama
· · Bribri (BZD)	5	Costa Rica
· Chibchan B		
· · San Blas Kuna (CUK)		Panama
· · Kogí (KOG)	4	Colombia
Misumalpan		
· Mískito (MIQ)	6	Nicaragua
· Mayangna (YAN)	4	Nicaragua
Camsá (KBH)	4	Colombia

TABLE 7.22

Some languages of Western Amazonia

Name	Size	Country
Chapacuran		
· Wari' (PAV)	4	Brazil
Guajiboan		
· Guahibo (GUH)	5	Colombia
Arawakan		
· Yanesha' (AME)	4	Peru
· Parecís (PAB)	4	Brazil
· Palikúr (PLU)	4	Brazil
· Northern Arawakan		
· · Wapishana (WAP)	5	Guyana
· · Caribbean Arawakan		
· · · Arawak (ARW)	4	Guyana
· · · Taino (TNQ)	0	Caribbean

(continued)

TABLE 7.22

Continued

Name	Size	Country
· · · Wayuu (GUC)	6	Colombia
· · · Garifuna (CAB)	6	Honduras
· · Inland Northern Arawakan		
· · · Curripaco (KPC)	5	Colombia
· · · Piapoco (PIO)	4	Colombia
· Southern Arawakan		
· · Bolivia-Parana		
· · · Terêna (TER)	5	Brazil
· · · Moxo		
· · · · Ignaciano (IGN)	4	Bolivia
· · · · Trinitario (TRN)	4	Bolivia
· · Purus		
· · · Apurinã (APU)	4	Brazil
· · · Yine (PIB)	4	Peru
· · Pre-Andine		
· · · Nomatsiguenga (NOT)	4	Peru
· · · Asha-Ashe-Kak-Matsi-Nan		
· · · · Machiguenga (MCB)	4	Peru
· · · · Asha-Ashe		
· · · · · Asháninka (CNI)	5	Peru
· · · · · Ashéninka		
· · · · · · Ashéninka Pajonal (CJO)	5	Peru
· · · · · · Apurucayali (CPC)	4	Peru
Arauan		
· Kulina (CUL)	4	Brazil
Harákmbut		
· Amarakaeri (AMR)	3	Peru
Puinavean		
· Puinave (PUI)	4	Colombia
Katukinan		
· Kanamarí (KNM)	4	Brazil
Tucanoan		
· Tucano (TUO)	4	Brazil
· Guanano (GVC)	3	Brazil
· Cubeo (CUB)	4	Colombia
Ticuna (TCA)	5	Brazil

have more than three or four. It has a particularly exuberant system of suffixes for expressing different degrees of past tense. [-hásij] expresses a proximate past, typically something that happened today; [-háj] is a recent past, but not today; [-sij] is a matter of weeks ago; [-tij] a matter of months; [-haⁿda] a matter of years (D. Payne 1985: 240–246).

TABLE 7.23
**Some languages of the northern foothills
of the Andes**

Name	Size	Country
Pumé (YAE)	4	Venezuela
Cofán (CON)	4	Colombia
Candoshi-Shapra (CBU)	4	Peru
Jivaroan		
· Jívaro		
· · Shuar (JIV)	5	Ecuador
· · Achuar-Shiwiar (ACU)	4	Ecuador
· · Huambisa (HUB)	4	Peru
· Aguaruna (AGR)	5	Peru
Cahuapanan		
· Chayahuita (CBT)	4	Peru
· Jebero (JEB)	4	Peru
Zaparoan		
· Iquito (IQU)	2	Peru
· Arabela (ARL)	2	Peru
· Záparo (ZRO)	1	Ecuador
Yagua (YAD)	6	Peru
Waorani (AUC)	4	Ecuador
Urarina (URA)	4	Peru
Boran		
· Bora (BOA)	4	Peru
Witotoan		
· Minica Huitoto (HTO)	4	Colombia
· Murui Huitoto (HUU)	4	Colombia
Andoque (ano)	3	Colombia

7.21 Andes region

Table 7.24 lists some of the languages of the Andes.

The Quechuan family (prefixed with a “p” in Figure 7.3) is spoken in Colombia, Ecuador, Peru, Bolivia, and Argentina. For more details on the Quechuan languages, see the sketch of Ayacucho Quechua (§7.27). Aymaran (prefixed with a “b” in Figure 7.3) has often been considered to be related to Quechuan, due in part to similarities in much of their vocabulary. But current thought is that the similar words are actually too similar and are not very common in the basic vocabulary: a state of affairs that suggests borrowing, not descent from a remote common ancestor.

Chipaya-Uru consists of two languages spoken in Bolivia (prefixed with a “c” in Figure 7.3). At last report, though, Uru had only two native speakers, so Chipaya is well on its way to becoming a language isolate.

TABLE 7.24

Some languages of the Andes

Name	Size	Country
Aymaran		
· Aymara (AYR)	7	Bolivia
· Jaqaru (JQR)	3	Peru
Quechuan		
· South Bolivian Quechua (QUH)	7	Bolivia
· Cusco Quechua (QUZ)	7	Peru
· Ayacucho Quechua (QUY)	7	Peru
· Chimborazo Highland Quichua (QUG)	6	Ecuador
Chipaya-Uru		
· Chipaya (CAP)	4	Bolivia
· Uru (URE)	1	Bolivia

7.22 Southern foothills

Table 7.25 lists some of the languages of the southern foothills of the Andes.

The features that get the most mention in descriptions of the languages of the southern foothills region are their agglutinative morphology, SOV word order, and ergative-absolutive morphosyntactic alignment. Panoan (plotted with a prefixed “^{ci}” in Figure 7.3) consists of 28 languages, of which 13 are no longer spoken. These languages are spoken in Bolivia, Peru, and Brazil. The Tacanan (prefixed “^e”) family consists of five languages spoken in the Amazonian lowlands of northern Bolivia and southeastern Peru. According to Kaufman (1990: 45), there is a very good chance that the Panoan and Tacanan families are genetically related.

7.23 South

In Table 7.26 we list several languages of the southern part of the continent, from Chile to the non-Amazonian part of Brazil. Indigenous language families of this region include Matacoan (prefixed with an “^f” in Figure 7.3), Guaykuran (prefixed “^g”), Mascoyan (prefixed “^m”), Zamucoan (prefixed “ⁿ”), and Jean (prefixed “^r”). In the Chaco province of Argentina, Wichí Lhamtés Vezoz is one of three local languages that share official status with Spanish. Since 1926, efforts have been made to group the Brazilian languages in this region into what has been called a *Macro-Jê* clade. But the evidence has been too spotty for any variant of the Macro-Jê hypothesis to gain firm acceptance.

Mapudungu contains two indigenous languages of Chile and Argentina, of which the most populous is Mapudungun. In many ways, Mapudungun clause structure is similar to that of Blackfoot (§7.3). There are no explicit markers on nouns to tell whether they are the subject or object of transitive verbs. Verbs, however, give some information about their subject and object. There are specific

TABLE 7.25

Some languages of the southern foothills of the Andes

Name	Size	Country
Yuracare (YUZ)	4	Bolivia
Panoan		
· Kaxararí (KTX)	3	Brazil
· Mainline Panoan		
· · Shipibo		
· · · Capanhua (KAQ)	2	Peru
· · · Marúbo (MZR)	4	Brazil
· · · Shipibo-Conibo (SHP)	5	Peru
· · · Panoan Katukína (KNT)	3	Brazil
· · Tri-State Panoan		
· · · Amahuaca (AMC)	3	Brazil
· · · Sharanahua (MCD)	4	Peru
· · · Yaminahua (YAA)	4	Peru
· · · Kashinawa (CBS)	4	Peru
· · Yora (MTS)	3	Peru
· · Cashibo-Cacataibo (CBR)	4	Peru
· Matsés (MCF)	4	Peru
· Chácobo (CAO)	3	Bolivia
Tacanan		
· Tacana (TNA)	4	Bolivia
· Cavineña (CAV)	4	Bolivia
· Ese Ejja (ESE)	3	Bolivia
Tsimané (CAS)	4	Bolivia

markers that tell they are first, second, or third person, so that a proposition like ‘I call you’ can be unambiguously expressed in the verb itself. In the common situation where both roles are filled by third-person words, the verb has different suffixes depending on whether the subject is the topic of the sentence. The topic usually comes first in the sentence. If the topic noun is the subject, a direct suffix appears on the verb (10a); if it is the object, then an inverse suffix is used (10b).

(10) Mapudungun

- a. wentʃu miʃsim-fi domo
 man called-DIR woman
 ‘The man called the woman.’ (Zúñiga 2006: 231)
- b. domo miʃsim-ewej wentʃu
 woman called-INV man
 ‘The man called the woman.’ (Zúñiga 2006: 233)
- c. ʔji ʔjaw ni ʔuka
 DEF.ART father his house
 ‘the father’s house’ (Zúñiga 2006: 100)

TABLE 7.26

Some languages of southern South America

Name	Size	Country
Qawasqar (ALC)	2	Chile
Mapudungu		
· Mapudungun (ARN)	6	Chile
Matacoan		
· Iyo'wujwa Chorote (CRQ)	4	Argentina
· Nivaclé (CAG)	5	Paraguay
· Wichi Lhamtés Vejoz (WLV)	5	Argentina
Guaykuruan		
· Mocoví (MOC)	4	Argentina
· Pilagá (PLG)	4	Argentina
· Toba (TOB)	5	Argentina
Mascoyan		
· Enlhet (ENL)	4	Paraguay
· Toba-Mascoy (TMF)	4	Paraguay
Zamucoan		
· Ayoreo (AYO)	4	Paraguay
· Chamacoco (CEG)	3	Paraguay
Chiquitano (CAX)	4	Bolivia
Bororoan		
· Borôro (BOR)	4	Brazil
Rikbaktsa (RKB)	2	Brazil
Jean		
· Xokleng (XOK)	3	Brazil
· Central Jean		
· · Xavánte (XAV)	4	Brazil
· · Xerénte (XER)	4	Brazil
Maxakalí (MBL)	4	Brazil
Iatê (FUN)	4	Brazil
Karajá (KPI)	4	Brazil
Ofayé (OPY)	1	Brazil
Guató (GTA)	1	Brazil

The fact that morphemes within the verb identify its subject, object, and other grammatical relations is an instance of HEAD MARKING, because the verb is the head of the clause. The opposite approach, DEPENDENT MARKING, is when the subject and object – dependents of the verb – have case markers to indicate their grammatical relation with the verb. Another construction that illustrates the difference between these two marking systems is when one noun is possessed by another. In the English phrase *the father's house*, *house* is the head and *the father's* is a dependent. It is the dependent that carries a marker of possession: dependent marking. The system works the opposite way in Mapudungun, where that

phrase would be translated as, literally, ‘the father his house.’ In this language it is the head noun that is preceded by a marker, [ɲi], that says the head is possessed by some third-person object: head marking (10c). The Americas have more than their share of languages that mostly use head marking on verbs; only Australia and New Guinea have comparable concentrations of such languages.

7.24 Central Amazonia

Indigenous languages of Central Amazonia (Table 7.27) include several isolates like Pirahã and Trumai, and the Jabutian (prefixed with a “k” in Figure 7.3), Nambiquaran (“l”), and Tupian (“m”) families.

TABLE 7.27
Some languages of Central Amazonia

Name	Size	Country
Pirahã (MYP)	3	Brazil
Itonama (ITO)	1	Bolivia
Kanoé (KXO)	1	Brazil
Jabutian		
· Arikapú (ARK)	1	Brazil
· Jabutí (JBT)	1	Brazil
Aikanã (TBA)	3	Brazil
Nambiquaran		
· Sabanê (SAE)	1	Brazil
· Southern Nambikuára (NAB)	3	Brazil
Irántxe (IRN)	2	Brazil
Trumai (TPY)	2	Brazil
Movima (MZP)	4	Bolivia
Tupian		
· Tupí-Guaraní		
· · Guaraní (GRN)		
· · · Eastern Bolivian Guaraní (GUI)	5	Bolivia
· · · Mbyá Guaraní (GUN)	5	Paraguay
· · · Paraguayan Guaraní (GUG)	7	Paraguay
· · · Kaiwá (KKG)	5	Brazil
· · · Nandeva (TPI)	4	Paraguay
· · Guayayú (GYR)	4	Bolivia
· · Tupí (TPW)		
· · · Cocama-Cocamilla (COD)	3	Peru
· · · Nheengatú (YRL)	5	Brazil
· · Tenetehara		
· · · Guajajára (GUB)	5	Brazil

(continued)

TABLE 7.27

Continued

Name	Size	Country
· · Kayabí		
· · · Kayabí (KYZ)	4	Brazil
· · Kawahib		
· · · Tenharim (PAH)	3	Brazil
· · Wayampí		
· · · Kaapor (URB)	3	Brazil
· · · Wayampí (OYM)	4	Brazil
· · Kamayurá (KAY)	3	Brazil
· Arikem		
· · Karitiána (KTN)	3	Brazil
· Awetí (AWE)	3	Brazil
· Jurúna (JUR)	3	Brazil
· Mondé (MND)	1	Brazil
· Mundurukú (MYU)	4	Brazil
· Ramarama		
· · Karo (ARR)	3	Brazil
· Sateré-Mawé (MAV)	4	Brazil
· Tuparí (TPR)	3	Brazil

Pirahã has garnered much attention, even in the popular press (Colapinto 2007), for claims that it is unusually simple. Everett (2005) pointed out, for example, that it has no true color terms – just words for ‘light’ and ‘dark’. He reported that it completely lacks any treatment of number: no grammatical marking for singular or plural, no quantifiers like ‘all’, and no number words, not even words for ‘one’ or ‘two’. Most impressively, he claims that the grammar lacks RECURSION – phrases that contain phrases of the same type – or any embedding at all. For example, one cannot construct sentences like ‘John thinks Mary caught the fish’, because that would constitute embedding a clause within another clause. Even constructions like ‘big slippery fish’ or ‘John’s father’s boat’ are impossible, and the language has no conjunctions. Controversies abound as to whether these claims about the language are true and, if so, whether the lack of recursion invalidates the idea that recursion is the only uniquely human property of human language (Hauser et al. 2002). Also controversial is Everett’s claim that these properties of the Pirahã language are entailed by the Pirahã culture in a sort of reverse linguistic relativity. He believes the Pirahã eschew abstractions and only talk about things in the personal experience of themselves and their interlocutors. Respondents (see, for example, comments published with Everett 2005) are not all convinced that such cultural traits, even if true, would necessarily lead to lack of recursion.

Old Guaraní or *Classical Guaraní* was widely spoken in the coastal regions of Brazil during that country’s early colonial period (1604–1767), and a variety of it served as a lingua franca among speakers of indigenous languages and colonizers

across Brazil. We know quite a bit about this language, thanks to grammars and dictionaries by colonial missionaries. In Paraguay, Guaraní has the status of an official language, alongside Spanish. Its nasal harmony rules are even more vigorous than those of Cubeo (§7.19) in several respects. Whereas in Cubeo, nasality only spreads perseveratively from a nasal syllable up to the next voiceless phoneme, in Guaraní it spreads in both directions, not stopping until another stressed syllable is encountered (11a; 2nd line shows underlying form). Nasality even spreads anticipatorily from voiced oral stops, which in Guaraní are prenasalized (11b).

- (11) Guaraní
- a. [ʔĩɲã.kãřãkú]
/iʝa.kãra'ku/
'is hot-headed'
- b. [řõⁿbowa'ta]
/ro-ⁿbo-wa'ta/
1SG.2SG-CAUS-walk
'I made you walk.' (both from Walker 1998: 230)

7.25 Northern Amazonia

Table 7.28 enumerates some of the languages and language families of Northern Amazonia.

The Cariban family (prefixed with an “n” in Figure 7.3) consists of 43 languages, of which 19 are no longer spoken. This family is found in Venezuela, Colombia, Suriname, Guyana, French Guiana, and Brazil. There are many subgroups in this family that do not appear to group together into major divisions. Hixkaryána is famous among linguists because of its rare basic word order: OVS. In (12), the verbal prefix [j-] means that both the subject and object are third person; thus it offers no help in deciding which of the nouns is subject and which is object. Only by invoking the default word order of the language is it clear that the boy is the catcher and not the catchee. Derbyshire (1985) discussed the significance of the existence of OVS word order for linguistic typology.

- (12) Hixkaryána
- kana j-anũm-nõ burjekõmõ
fish 3SG-catch-REC boy
'The boy has caught a fish.' (Derbyshire & Pullum 1981: 194)
REC = RECENT PAST.

The Island Carib (CRB) language, which used to be spoken on a number of Caribbean islands, belonged to the Arawakan language family, not Cariban. It is most closely related to the Central American language Garifuna CAB (§7.19). It

TABLE 7.28

Some languages of Northern Amazonia

Name	Size	Country
Cariban		
· Carib (CAR)	4	Venezuela
· North Amazonian Cariban		
· · Macushi (MBC)	5	Brazil
· Pemon (AOC)	4	Venezuela
· · Kapong		
· · · Akawaio (AKE)	4	Guyana
· · · Patamona (PBC)	4	Guyana
· South Amazonian Cariban		
· · Eñapa Woromaipu (PBH)	4	Venezuela
· Waiwai		
· · Waiwai (WAW)	4	Brazil
· · Hixkaryána (HIX)	3	Brazil
· Central Cariban		
· · Wayana (WAY)	3	Suriname
· · Maquiritari (MCH)	4	Venezuela
· Yukpa		
· · Yukpa (YUP)	4	Colombia
Yanomaman		
· Yanomamö (GUU)	5	Venezuela
Warao (WBA)	5	Venezuela
Sálivan		
· Piaroa (PID)	5	Venezuela
· Sáliba (SLC)	4	Colombia
Arutani (ATX)	2	Brazil
Sapé (SPC)	1	Venezuela
Yuwana (YAU)	3	Venezuela

appears to have got its name because Carib invaders from the mainland of South America imposed their nationality and name on their Arawakan subjects, even though they were unable to replace the native Arawakan language of their subjects.

The Yanomaman family consists of four languages found in Venezuela and Brazil. The Sálivan family (prefixed with an “^{co}” in Figure 7.3) consists of two languages of Venezuela and Colombia.

7.26 Sketch of Central Alaskan Yup'ik

7.26.1 GENETIC AFFILIATION AND GENERAL INFORMATION

Central Alaskan Yup'ik is a Yupik language belonging to the Eskimo branch of the Eskimo-Aleut family of languages. Yupik languages are spoken in both Siberia (the Chuckchi Peninsula) and Alaska, which makes Eskimo-Aleut the only family

that is definitively known to have been found in both North America and Asia in pre-Columbian times. Yup'ik is the largest of the Yupik languages, and the only one that we know is still being passed on naturally to children in a substantial number of villages. It stretches from Nunivak Island, along the Alaska coast from Bristol Bay to Norton Sound, and inland along the Nushagak, Kuskokwim, and Yukon Rivers. Its closest relative is Pacific Gulf Yupik. A somewhat more distant, Siberian, group of Yupik languages is spoken on the Siberian side of the Bering Sea, but Central Siberian Yupik, despite its name, is spoken more on St. Lawrence Island in Alaska. The following sketch of Yup'ik Eskimo is based mainly on Reed (1977).

Yup'ik is the only Yupik language in which the [p] in its name is long, which is indicated in the standard orthography by the apostrophe; the names of the other languages, as well as the name of the Yupik group in general, do not take the apostrophe.

Eskimo languages spoken in Alaska have been influenced by Russian and have borrowed extensively from that language:

- (13) a. [sta:lɪsta], [əsta:lɪsta] < Russian ['starasta] 'church elder'
 b. [tupu:lɪq] < Russian [ta'por] 'axe'
 c. [ku:lɪtʃaq] < Russian ['kur'itsa] 'chicken'

In Canada, and in recent years in Alaska, there has been more borrowing from English:

- (14) a. [pəluməsaq] 'women's underpants' < English *bloomers*
 b. [minaq] < English *miner*
 c. [ti:vi:q] < English *TV*

Kalaallisut Eskimo in Greenland has borrowed words mainly from Danish.

7.26.2 PHONETICS, PHONOLOGY, AND ORSTHOGRAPHY

Although a native speaker of Yup'ik invented a syllabography for the language around 1900, Yup'ik is now written in the Latin script using an orthography designed by American linguists. Table 7.29 gives the Yup'ik consonants in IPA; the standard orthography, where it differs, is in angled brackets.

There are long consonants, which are marked by an apostrophe in the standard orthography, as in the name of the language, <Yup'ik> [jup'ik]. Voiced fricatives are found only between voiced sounds. Because there is no need to indicate any voicing contrast in other positions, the simpler, single graphemes stand for voiceless fricatives there.

[tʃ] has the allophone [tʃ̥] before [ə]. [v] has the allophone [w] between most vowels.

Yup'ik has a triangular vowel system: high [i], [u], low central [a], and [ə], which is spelled <e>. All vowels except [ə] may occur long, which is indicated in the

TABLE 7.29

Yup'ik consonants

Stop	p	t	ṭ̥ ⟨c̥⟩	k	q
Fricative	f ⟨vv⟩	s ⟨ss⟩		x ⟨gg⟩, x ^w ⟨w⟩	χ ⟨rr⟩
· · lateral		ṭ̥ ⟨ll̥⟩			
· · +voice	v	z ⟨s⟩		ɣ ⟨g⟩, ɣ ^w ⟨ug̃⟩	ʁ ⟨r⟩, ʁ ^w ⟨ur̃⟩
· · lateral		ḡ ⟨l̥⟩			
Nasal	m	n		ŋ ⟨ng⟩	
· · -voice	m̥ ⟨m̥⟩	n̥ ⟨n̥⟩		ŋ̥ ⟨n̥g̥⟩	
Approximant		l	j ⟨y⟩		

orthography by gemination. [i] and [u], short or long, have lowered allophones [e] and [o] next to uvular consonants.

Stress in Yup'ik is predictable, and therefore is not indicated in the spelling. In many phonetic environments, depending in part on stress, sounds must be long, and in those cases their length isn't marked in the spelling either. The system is complicated, but can mostly be reduced to the following ordered rules:

- (15) a. Divide the word into syllables, splitting adjacent consonants and assigning intervocalic consonants to the syllable of the second vowel.
- b. Stress any syllable with a HEAVY NUCLEUS (diphthong or long vowel).
- c. Stress the initial syllable if it is closed (ends in a consonant).
- d. Working from left to right through the word, if two unstressed syllables come together, stress the second one; but if the first is closed and the second is open, stress the first one.
- e. Lengthen any short consonant after a stressed [ə].
- f. Lengthen any short consonant that has an unstressed short vowel on its left and a heavy nucleus on its right.
- g. Lengthen any short vowel in an open stressed syllable.
- h. Stress any syllable that precedes a syllable with a heavy nucleus.

The orthography also omits marking voicelessness when that feature is required by the phonology. Recall that fricatives are fundamentally written with double letters (e.g. ⟨vv⟩ [f] versus ⟨v⟩ [v]). But next to a stop or at the end of a word, fricatives must be voiceless, and the orthography calls for the shorter spelling (e.g. ⟨v⟩ spells [f] in places where [v] is impossible). Nasal consonants are always voiceless after voiceless obstruents, in which case the orthography omits the diacritic that marks nasal voicelessness in other environments.

7.26.3 MORPHOLOGY

Yup'ik has the very rich morphology of a typical polysynthetic language. Nouns and verbs begin with a root, which may often be followed by several lexical suffixes before an inflectional ending. Often enclitics are added after the ending.

- (16) aŋja-ɬi:-tʃiq-suɣnaχ-quq=ɬu
 boat-make-FUT-probably-3SG.V_{INTR}=also
 ‘Also, he will probably make a boat.’

Here, [aŋja] is the base, [ɬi], [tʃiq], and [suɣnaχ] are lexical suffixes, [quq] is the inflectional ending, and [ɬu] is an enclitic.

It should be noted that Yup’ik has very complicated morphophonemics, which cause almost every morpheme in the language to have many allomorphs.

7.26.3.1 Nouns

Nouns in Yup’ik are inflected for case, number, and possession. The case endings will be discussed below (see Table 7.30). There are three numbers: singular, dual, and plural. Sometimes the case, number, and possessor markers are so fused together that it is difficult, if not impossible, to say where one ends and the other begins. In any case, all three constitute part of the ending.

7.26.3.2 Case endings

ABSOLUTIVE

This case is used for the subject of an intransitive verb and the object of a transitive verb:

- (17) a. aŋja: tak:uq
 his.boat[abs] be_long
 S V_{INTR}
 ‘His boat is long.’
- b. aŋja: taŋχa:t
 his.boat[abs] they_see
 O V_{TR}
 ‘They see his boat.’

TABLE 7.30
Yup’ik case endings

Case	Abbrev	SG	DU	PL
Absolutive	ABS	—	-k	-t
Ergative	ERG	-m	-k	-t
Ablative	ABL	-mæk	-ɣnək	-nək
Locative	LOC	-mi	-ɣni	-ni
Perlative	PER	-kun	-ɣnəɣun	-txun
Essive	ESS	-tun	-xtun	-tʃətun
Allative	ALL	-mun	-ɣnun	-nun

ERGATIVE

Note that the plural and dual forms of this case are identical to the absolutive case forms.

This case marks the subject of transitive verbs, as in many languages. But it can also act like a genitive case, marking, for example, possession:

- (18) a. aŋu:tə-m nəʃ-a: nəqa
man-ERG eat-TR.3SG>3SG fish[abs]

‘The man is eating the fish.’

The symbol > shows which of two adjacent agreement markers is for the subject – the one at the open end of the symbol, here the left – and which is for the object.

- b. aŋu:tə-m qimuxt-i: nəʃ-uq
man-ERG dog-his eat-INTR.3SG

‘The man’s dog is eating.’

In (18b), possession is marked both on the dependent (ergative case on ‘man’) and on the head (possessive suffix on ‘dog’). Note also that because no direct object is present, the verb ‘eat’ has an intransitive ending.

ABLATIVE

This case suffix marks the point of origin (‘from’). It also takes on many functions of an instrumental case. In addition, some verbs of giving use the absolutive to mark the person to whom something is given, and they mark the object given by a noun with the ablative case. The case may also mark the subject matter of speaking and thinking, like the English preposition ‘about’.

- (19) təyʉʉχua nu:si-ni əstu:ʒu-mək
he_took knife-his table-ABL

‘He took his knife from the table.’

The ablative may also be used for indefinite objects (20a). Note that the verbs in such sentences are marked as intransitive, because the object is not an actual argument of the verb, which would be absolutive. Compare the marking on verbs that take a definite object (20b).

- (20) a. taŋəχ-tuq iməʃmiutaʃ-mək
see-INTR.3SG mink-ABL

‘He sees a mink.’

- b. taŋəχ-a: iməʃmiutaq
see-TR.3SG>3SG mink[abs]

‘He sees the mink.’

LOCATIVE

This case suffix marks location (21a) and the base of comparison (21b):

- (21) a. nəʃ:-uq nə-m:i
eat-INTR.3SG house-LOC
'He is eating in the house.'
- b. una əna asi-nəu-uq nə-vni
this house[abs] be_good-COMP-INTR.3SG house-2SG.LOC
'This house is better than yours.'

PERLATIVE

The PERLATIVE CASE (PER) marks the route of motion. It can also mark the instrument by which an action is performed. It also marks the partitive, that is, the object when only a part of a whole is affected.

- (22) a. ajax-tuq kui-pax-kun
go-INTR.3SG river-big-PER
'He is going by way of the Yukon River.'
- b. navγ-a: aŋja-ni kiŋ:ua-kun
break-TR.3SG>3SG boat-his[abs] stern-PER
'He broke his boat on its stern.'
- c. aŋja-mi-kun tək:i:təʎ-u:q
boat-3SG.POSS-PER arrive-INTR.3SG
'He arrived using his boat.'

ESSIVE

This case indicates similarity and is usually translated by the English prepositions 'like' and 'as'. It is also used to name the language of something or at what price one is selling something.

- (23) a. kas:aq jup:iχ-tun juʃaχ-tuq
white_man[abs] Yup'ik-ESS dance-INTR.3SG
'The white man is dancing like a Yup'ik.'
- b. tʃamək atə-ŋqəχ-ta tauna jup:iχ-tun
what name-have-be that Yup'ik-ESS
'What is the name of that in Yup'ik?'

ALLATIVE

This case ending marks motion to or into something.

- (24) aʃna-m əkəʎχu-a nu:ziq ja:si:γ-mun
woman-ERG put-TR.3SG>3SG knife[abs] box-ALL
'The woman put the knife into the box.'

7.26.3.3 Possessive suffixes

Besides being inflected for case, nouns can also be inflected for possession. For the sake of keeping things simple and brief, only a few examples of possessive endings will be shown here.

- (25) a. [nuna] 'land' ABS
 b. [nunaa] 'his land' ABS < [nuna-ŋa]
 c. [nuni:n] 'his land' ERG
 d. [uiŋa] 'her husband' ABS
 e. [uiŋan] 'her husband' ERG
 f. [uiŋit] 'their husbands' ABS

7.26.3.4 Pronouns

Because the person and number of the subject and object of the verb are both marked on the verbs, personal pronouns are not used as much in Yup'ik as they are in English. Yup'ik pronouns and person markers do not distinguish gender but do distinguish three numbers: singular, dual, and plural. Like nouns, pronouns are also inflected for various cases. The various forms will not be listed here.

Unlike English, Yup'ik has different third person markers depending on whether or not they refer to the subject of the preceding clause. Those that refer to the same object are COREFERENTIAL PRONOUNS. In English, a sentence like 'Doris is happy because she kissed Tom' is ambiguous because the pronoun 'she' could be referring to Doris or to someone mentioned earlier. In Yup'ik there is no such ambiguity (T. Payne 1980).

Yup'ik has demonstratives with several unusual features. In addition to the familiar distinction of proximal versus distal, demonstratives can also be EXTENDED. This last refers to things that are spread out horizontally. Demonstratives also can locate things by many other spatial characteristics, including their relation to the speaker or the hearer; whether it is up or down, in or out; and, most robustly, its relation to the river: toward the river, away from the river, across the river, upstream, downstream. For most of these spatial characteristics, there are distal, proximal, and extended variants, resulting in a system of some 30 different demonstrative stems (Miyaoaka 2012: 346).

7.26.3.5 Verbs

Yup'ik verbs take endings that mark transitivity and intransitivity, person and number of the subject and object, and mood. In some cases, it is possible to observe the sequencing of the various markers in verbal endings:

- (26) a. taŋχ-a-i-t
 see-TR-3PL.OBJ-3PL.SBJ
 'They see them.'

But at other times, it is not possible to isolate the various elements, evincing a strong tendency toward fusion:

- (27) a. nał:u-uq
not_know-INTR.3SG
'He doesn't know.'
- b. nał:u-a
not_know-TR.3SG>3SG
'He doesn't know it.'

It is common in Yup'ik to have verbs that are derived from nouns. Even though such verbs need to be translated into English as a transitive verb phrase, where the base noun is expressed as a separate word that is a direct object, they are basically intransitive in Yup'ik and take intransitive endings. Thus (28a) is marked as an intransitive verb with an implicit third person singular subject; there is no transitive ending referring to the boat. If such verbs are marked as transitive, it is to show that they are taking an additional argument, such as a beneficiary (28b).

- (28) a. anja-ḡ:i-uql
boat-make-INTR.3SG
'He is making a boat.'
- b. anja-ḡ:i-a
boat-make-TR.3SG>3SG
'He is making a boat for him.'

MOODS

There are many mood forms in Yup'ik. The indicative is unmarked. It is used in declarative sentences and polar questions, but content questions must use verbs marked for a *INTERROGATIVE MOOD*. The imperative mood, as in Hawaiian, can be used not only for second-person commands, but also for wishes, suggestions, and so forth in the first and third persons as well.

The *SUBORDINATE MOOD* involves a variety of participle-like constructions or dependent clauses, which can be translated into English by subordinate clauses such as 'while I was doing X', 'when I was doing X', or 'before I did X'.

- (29) a. [tʃaḡi:ŋinauɓnani] 'while he works'
b. [tʃaḡi:vaiḡyan] 'before he works'

7.26.4 SYNTAX

Thanks to all the grammatical work done by the morphology, Eskimo syntax is relatively simple. Grammatical relations are marked primarily by case endings.

The basic word order is SVO, although examples of OV order are commonly found if there is no subject noun in the sentence. As can be seen from the discussion of case endings above, Yup'ik is an ergative-absolutive language.

7.26.4.1 Negation

Negation is expressed primarily by various suffixes. In addition, some verbs have very different roots for their negative counterparts. For example, there are different stems for 'to know' [nałunxitə-] and 'not to know' [nału-].

7.26.4.2 Interrogative sentences

Polar questions add the interrogative enclitic [=qa:] to the first word in the sentence:

- (30) aŋu:t-əm=qa: nəʔ-a: nəqa
 man-ERG=Q man-TR.3SG>3SG fish[abs]
 'Is the man eating the fish?'

Content questions inflect their main verb for the interrogative mood:

- (31) a. tʃi:n nəʔ:-a
 why eat-INTR.3SG.Q
 'Why is he eating?'
 b. tʃi:n nəʔ-au
 why eat-TR.3SG>3SG.Q
 'Why is he eating it?'

7.26.4.3 Relative clauses

Yup'ik relative clauses are formed by various suffixes, which basically nominalize the verbs to which they are attached. The suffix [-ʔ(ə)χ-] with no possessive ending forms a noun taken to be the agent of the verbal action:

- (32) pai-ʔəq
 babysit-REL[abs]
 'the one who babysat'

But if a possessive ending is added, the word denotes the patient of the verbal action, with the possessor as the agent:

- (33) pai-ʔχ-a
 babysit-REL-3SG.POSS
 'the one she babysat'

7.26.5 SAMPLE TEXT

Many Yup'ik texts are translations from English done by linguists and their native-language consultants. The text given here is a very simple self-introduction that a native teacher of Yup'ik gave to Lyovin in St. Mary's, Alaska, in 1978, when he was training some Yup'ik native speakers to become elementary school teachers of that language.

- (34) Waqaa! Wiinga atengqertua Lena L. . . amek.

x^waq:a: x^wi:ŋa atə-ŋqəχ-tu-a ɬəna ɬ. . .-amək
 hello 1SG name-have-INTR-1SG Lena L. . .-ABL
 'Hello! My name is Lena L.'

- (35) Yup'iunga. Yugtun qaneryuumaunga.

ju-p'i-u-ŋa jux-tun qanəχ-ju:ma-u-ŋa
 person-real-INTR-1SG Yup'ik-ESS speak-can-INTR-1SG
 'I am Yup'ik. I can speak Yup'ik.'

- (36) Elicariciqua mikelnguarnek yugtun.

əɬi:tʃaχ-i-tʃiq-u-a mikəɬ-ŋuaχ-nək jux-tun
 teach-INTR-FUT-INTR-1SG small-one-ABL.PL Yup'ik-ESS
 'I'll be teaching Yup'ik to children.'

[əɬi:tʃaχ-i-tʃiq-u-a]. [-i-] is a detransitivizer, turning a transitive verb into an intransitive one suitable for use with indefinite objects.

- (37) Yuurtellruunga nunacuarmi atengqerrluni Anipaunguarvigmek.

ju:χtə-ɬχu-u-ŋa nuna-tʃuaχ-mi atə-ŋqəχ-ɬu-ni
 be_born-PFV-INTR-1SG village-small-LOC.SG name-have-APPOSITIONAL.INTR-3SG
 ani:pa-uŋuaχ-viχ-mək
 snowy_owl-like-place-ABL.SG

'I was born in a small village named Anipaunguarvik'

- (38) Elicartua qayugga elicallerkamnek mikelnguarat yugtun.

əɬi:tʃaχ-tu-a qaju:xa əɬi:tʃa-ɬəχ-kam-nək mikəɬ-ŋuaχa-t jux-tun
 learn-INTR-1SG how teach-NOUN-FUT-ABL.PL small-one-Abs.PL Yup'ik-ESS
 'I am learning how to teach the children Yup'ik'

7.27 Sketch of Ayacucho Quechua

7.27.1 GENETIC RELATIONSHIP AND GENERAL INFORMATION

Quechua is yet another example of a language clade that people view with a wide range of perspectives. At one extreme, it is a language isolate (QUE) that is spoken over a large region: from Putumayo in southern Colombia to Santiago del Estero in northern

Argentina – the approximate extent of the Inca Empire. Perhaps the idea that Quechua constitutes a single language dates back to the Incas, who made the Quechuan dialect of its homeland and capital, Cusco, the official language of their empire. Even today, the Academia Mayor de la Lengua Quechua insists that the language of the city of Cusco should be the single authoritative standard for all of Quechua. At the other extreme, Ethnologue treats it as a family containing 45 languages. Ayacucho Quechua, for example, has been assigned the distinct language code QUY. There is certainly a great deal of variation and lack of mutual intelligibility within Quechua.

The Ayacucho variety of Quechua, sometimes called Chanka, is the variety we concentrate on in this sketch. It is closely related to the language of Cusco and is fairly close to other Quechua dialects in southern Peru, Bolivia, Chile, and Argentina. These varieties are often considered as constituting a reasonably cohesive language called Southern Quechua, or, following its self-designation, *Qhichwa* [q^het̪wa]. With perhaps 7 million speakers, Southern Quechua would be the largest indigenous language in the Americas, and one of very few surviving indigenous languages in Argentina. The group most similar to Southern Quechua are some dozen dialects spoken by over a million people in the north: throughout Ecuador, reaching north into Putumayo and south into the northern third of Peru. These northern dialects are often considered as constituting a single language called *Kichwa*, whose standard form is based primarily on Chimborazo Highland *Quichua*, which is spoken by 1 million speakers in Ecuador. *Kichwa* has a somewhat simpler morphology than Southern Quechua and some significant sound changes, including one that changed all uvular stops like [q^h] to [k] – hence the distinction in the language names. Perhaps unexpectedly, the northern and southern dialects are much more similar to each other than either are to the Quechua languages of central Peru situated between them. Consequently Southern Quechua and *Kichwa* form a group called *Peripheral Quechua*. The Central Quechua dialects are not only very different from *Peripheral Quechua* but are quite different among themselves as well. This linguistic situation suggests that the *urheimat* of Quechua was in central Peru. No doubt different dialects evolved there for many centuries until a particular dialect quickly spread north and south in fairly recent times – likely as a consequence of the Inca expansion.

Estimates of the total number of speakers of some kind of Quechua vary widely, from 6 to 10 million. In Ecuador and Bolivia, over 20% of the population speaks Quechua, and sizeable populations are found also in Peru and several other countries of South America. Four of the largest Quechua-speaking groups of population are plotted with a prefixed “P” in Figure 7.3. In some areas there are still very large numbers of *MONOLINGUAL* Quechua speakers – people who speak only Quechua – and some Indians are learning Quechua as a second language. Some groups of Quechua speakers are reported to be trilingual in Quechua, Spanish, and Aymara. Most Quechua speakers are to a certain extent bilingual in Spanish, and the latter language has greatly influenced Quechua. In both Cusco and Ayacucho Quechua, there are not only many basic vocabulary items borrowed from Spanish

but also some borrowed grammatical words such as prepositions, which originally did not exist in Quechua as a grammatical category. One such borrowed preposition is [asta] ‘until’ (< Spanish *hasta*), which is optionally used with nouns to which the Quechua suffix [-kama] ‘until’ is attached. The Quechua mediopassive construction (§7.27.3.4) is sometimes used like the Spanish passive, most likely because of Spanish influence.

7.27.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

7.27.2.1 Consonants

The consonants found in the native vocabulary of Ayacucho Quechua are shown in Table 7.31. In addition, voiced stops ([b], [d], [g]) are found in loans from Spanish, as is [f] and the fricative [z], which is substituted for the trilled [r] of Spanish. The symbols in angled brackets reflect modern Peruvian orthographic standards (1985), which in recent decades have veered away from Spanish-based standards.

Other dialects, including nearby Cusco, include ejectives and aspirated stops. Those features are indicated in modern spelling by placing an apostrophe ‘>’ and an <h> after the consonant letter, respectively.

The voiceless uvular fricative phoneme [χ] corresponds to a voiceless uvular stop [q] in other Southern Quechuan dialects; in the standard orthography, the symbol <q> is used to maximize orthographic uniformity between the dialects. Actually, even in Ayacucho Quechua, [q] may optionally be heard as an allophone of [χ] after [n]. The phoneme [h] varies in value from [h] to [x]. The phoneme [n] is realized as [ɲ] in the coda of a syllable except before alveolar or palatal obstruents, where it is realized as [n].

7.27.2.2 Vowels

Quechua has a triangular three-vowel system, [i], [u], and [a]. The high vowels are lowered to mid position ([e] and [o]) when next to a uvular consonant. If only native Quechua words are considered, this variation would clearly be considered allophonic and therefore should not be indicated in a phonemic transcription or in orthography, which traditionally aims for phonemic levels of representation. The

TABLE 7.31
Quechua consonants

Stops	p	t	ʈʂ <ch>	k	
Fricatives		s		χ <q>	h
Nasals	m	n	ɲ <ñ>		
Laterals		l	ʎ <ll>		
Tap		r <r>			
Glides			j (y)	w	

analysis is muddled somewhat by the fact that Quechua has multitudinous loans from Spanish, and many speakers, especially those familiar with Spanish, render Spanish [e] and [o] without change: e.g. [enero] ‘January’. For registers of Quechua where [e] and [o] consequently appear in many words even when not next to a uvular, there are five different vowel phonemes. Some orthoepists, therefore, prefer spelling Quechua with five vowels. The modern trend, however, is to spell Quechua with just three vowels, even loanwords like [iniru], ignoring the fact that many people attempt a more Spanish pronunciation of the vowels in such words. In the following orthographic transcriptions, we normally spell loanwords in their fully nativized, three-vowel form in the orthography, but if the source of an example gives the word with Spanish phonemes, these are indicated in the IPA transcription.

Stress is generally placed on the penultimate syllable, but in a few words the stress falls on the final syllable.

7.27.3 MORPHOLOGY

Quechua morphology is agglutinative in the sense that relatively long strings of suffixes are common. At the same time, a number of morphophonemic irregularities make it atypical of agglutinative languages. The degree of irregularity, however, is not so high as to qualify Quechua as a polysynthetic language according to our definition (§1.2.1).

7.27.3.1 Nouns

Inflectional suffixes of a Quechua noun are ordered as follows:

(39) Noun stem (+ person suffix) (+ plural suffix) (+ case suffix)

Many stems can be used either as noun or verb stems.

PERSONAL SUFFIXES

Personal suffixes on nouns have the force of possessive pronouns. On nominalized verb forms, they indicate the subject of the original verb.

(40)

1	-ni-	
1.INCL	-ni-ntʃik	
2	-ni-(j)ki	
3	-ni-n	

(41a) has basic examples. (41b) illustrates that if the noun ends in a vowel, the suffix [ni] is dropped. There is no differentiation as to gender, so the third person means either ‘she’ or ‘he’; but not ‘it’, because these suffixes apply only to people. The corresponding plurals are formed by adding to one of the above endings [-ku] or, in the second person, [-tʃik] (41c). Adding [-ku] to a first person form makes an exclusive plural. No plural suffix is added to the inclusive ending [-ni-ntʃik], which is inherently plural.

- (41) a. [ɲan] 'road': [ɲan-ni-j] 'my road', [ɲan-ni-ki] 'your road', [ɲan-ni-n] 'his road',
[ɲan-ni-ki-tʃik] 'your and my road'
- b. [wasi] 'house': [wasi-j] 'my house', [wasi-ki] 'your house', [wasi-n] 'her house',
[wasi-ntʃik] 'your and my house'
- c. [ɲan-ni-j-ku] 'our road', [wasi-j-ku] 'our house', [ɲan-nin-ku] 'their road',
[wasi-ki-tʃik] 'your house'

PLURAL MARKING

Number, except in personal suffixes and independent pronouns, is not an obligatory category in Quechua. Nouns are optionally pluralized by the addition of the suffix [-kuna] (42a). It can follow a personal suffix and even a plural suffix that applies to the personal suffix. Thus (42b) means 'the houses that belong to you all'.

- (42) a. [wasi] 'house' or 'houses'; [wasi-kuna] 'houses'
- b. [wasi-ki-tʃik-kuna] 'your houses'

CASE ENDINGS

Cases in Ayacucho Quechua are signaled primarily by suffixes on the nouns, as well as by a few prepositions borrowed from Spanish. The case suffixes are listed in Table 7.32.

The main grammatical cases – nominative, accusative, and genitive – work basically as you would expect them to work in a nominative-accusative language. The accusative is used for objects of verbs (43a); with verbs of motion, it shows the goal of a person's movement (43b). In addition, the accusative forms of adjectives serve as adverbs (43c). The genitive marks a possessor (43d). The possessed noun has to have a possessive suffix agreeing with the possessor – a combination of head and dependent marking. Thus, (43d) is literally 'of me my father's his house'.

TABLE 7.32

Case endings in Ayacucho Quechua

	Case	Suffix	Use
NOM	nominative	-∅	subject of verb
ACC	accusative	-ta	object, goal, adverbial
GEN	genitive	-pa	possession, noun adjunct
BEN	benefactive	-paχ	'for'
CAUSE	causal	-rajku	'because of'
LOC	locative	-pi	'at'
ALL	allative	-man	'to'
ABL	ablative	-manta	'from', 'than'
TERM	terminative	-kama	'up to'
INTER	interessive	-pura	'among'
INS	instrumental	-wan	'with', 'and'

- (43) a. wasi-ta χawa-n
 house-ACC watch-PRS[3]
 ‘He watches the house.’
- b. wasi-ta ri-n
 house-ACC go-PRS[3]
 ‘He goes to the house.’
- c. aχin-ta ruwa-n
 good-ACC do-PRS[3]
 ‘He does well.’
- d. juχa-pa tajta-j-pa wasi-n
 1SG-GEN father-1SG-GEN house-3SG
 ‘my father’s house’

The other cases are fairly well described by the English prepositions given in Table 7.32. The benefactive shows purpose (44a) as well as a beneficiary (44b). The causal case, which is not very common in the world’s languages, marks the reason for something.

- (44) a. miku-na-m-paχ
 eat-NOUN-3SG-BEN
 ‘in order for him to eat’
 Literally, ‘for his eating.’
- b. χam-λa-paχ=mi
 2SG-only-BEN=EXPER
 ‘It is just for you.’

The remaining cases express various types of location. The locative generally expresses a point in space or time, without the idea of movement.

- (45) a. wasi-pi
 house-LOC
 ‘in, at, or on the house’
- b. sitimbri-pi
 September-LOC
 ‘in September’

The allative indicates goal of motion. Unlike the accusative, it works with objects as well as people. It is sometimes called the *dative* case because it can show recipient (46b). With [hina], which by itself means ‘like’, it means ‘according to’ (46c). This is one of many instances in which locational cases are used to express ideas that aren’t locational.

- (46) a. maju-man riṭka-n-i
 river-ALL go-PRS-1SG
 ‘I am going to the river.’
- b. juḡa-man ḡumu-wa-j
 1SG-ALL give-1SG.OBJ-IMP
- c. paj-man hina
 3SG-ALL like
 ‘according to him’

The ablative marks the origin of an action; in addition, it has many idiomatic uses. It may be translated as ‘from’, ‘about’, ‘concerning’, ‘instead of’, ‘made of’, or ‘by’.

- (47) a. wasi-manta ḡuxsi-n
 house-ABL leave-PRS[3]
 ‘He leaves the house.’
- b. wasi-manta rima-n
 house-ABL speak-PRS[3]
 ‘He speaks about the house.’
- c. paj=mi ri-nḡa juḡa-manta
 3SG=EXPER go-FUT 1SG-ABL
 ‘He will go in my stead.’
 EXPER = EXPERIENTIAL enclitic, which indicates personal experience.
- d. hizu-manta=m
 iron-ABL=EXPER
 ‘It is made of iron.’

The INTERESSIVE CASE (INTER) is not very common among the languages of the world. It shows that something is located in the midst of other things of the same kind.

- (48) amigu-pura ka-tṭka-n
 friend-INTER be-DUR-PRS[3]
 ‘They are among friends.’

When the things involved are not of the same kind, the word [tṭawpi], meaning ‘center’ or ‘midst’, is placed in the locative case after the noun. Such use of nouns with a locational meaning is a common way Quechua expresses locations that are more precise than the available cases.

- (49) wasi tṭawpi-pi ka-tṭka-n
 house midst-LOC be-DUR-PRS[3]
 ‘He is among the houses.’

The instrumental case, in addition to expressing means (50a), can be used as a comitative case to show accompaniment (50b). This may remind you of the Russian instrumental case, which can also be used as a comitative. The decision as to what name to give cases that have multiple functions is ideally based on what their most basic, core function is. But, frankly, very often the choice is not clearcut, and different linguists may give different names to the same case in the same language.

- (50) a. *lampa-wan* *ʎamka-tʃka-n*
 hoe-INS work-DUR-PRS[3]
 'He is working with the hoe.'
- b. *ɲuχa-wan=mi ri-n*
 1SG-INS=EXPER go-PRS[3]
 'He goes with me.'

Some of the case suffixes may occur in combination with other case suffixes. For example, if a noun in the genitive case is used elliptically, omitting the head, its genitive ending is followed by the ending that would have appeared on the head noun:

- (51) *kura-pa-pi*
 priest-GEN-LOC
 'at the priest's (house)'

The instrumental case suffix, too, may follow all other suffixes, where it serves to connect two nouns, as a conjunction 'and':

- (52) *wasi-ta-wan tuzi-ta*
 house-ACC-INS tower-ACC
 'to the house and to the tower'

7.27.3.2 Pronouns

Independent personal pronouns are used less than in English. They would mostly be semantically redundant, because the persons of the subject and object are marked on most verb forms. They are used for emphasis and in situations in which the verbal suffix system has insufficient detail. The independent forms are shown in Table 7.33. Note how the plural is formed by adding, in an agglutinative

TABLE 7.33

Quechua personal pronouns

Person	SG		PL
1	[ɲuχa]	EXCL:	[ɲuχa-j-ku]
		INCL:	[ɲuχa-nʃik]
2	[χam]		[χam-kuna]
3	[paj]		[paj-kuna]

fashion, plural endings we have already encountered for nouns and their personal suffixes.

7.27.3.3 Adjectives

For the most part, Quechua adjectives behave very much like nouns; in fact some words may act either as nouns or as adjectives. The comparison of adjectives does not involve any specific comparative or superlative inflection; in this respect it is like Japanese and many other languages. Note the use of the ablative case to mark the basis of comparison.

- (53) a. $\chi\chi\alpha$ hina kapka=m kaj tanta= $\chi\alpha$
 rock like hard=FOC this bread=TOP
 ‘This bread is as hard as a rock.’
 FOC = FOCUS, which indicates new information.
- b. $\chi\chi\alpha$ -manta aswan kapka=m kaj tanta= $\chi\alpha$
 rock-ABL more hard=FOC this bread=TOP
 ‘This bread is harder than a rock.’
- c. λ ama-kuna-manta pisi hatun=mi a $\lambda\chi$ u-kuna= $\chi\alpha$
 llama-PL-ABL little big=FOC dog-PL=TOP
 ‘Dogs are smaller than llamas.’
- d. kaj sat \bar{t} a= $\chi\alpha$ λ iw-manta aswan hatun-in=mi
 this tree=TOP every-ABL more big-3=FOC
 ‘This tree is the biggest of all.’
- e. kaj sat \bar{t} a= $\chi\alpha$ λ iw sat \bar{t} a-kuna-manta pisi hatun=mi
 this tree=TOP every tree-PL-ABL little big=FOC
 ‘This tree is the smallest of all trees.’

7.27.3.4 Verbs

Quechua verbs inflect for many grammatical categories, including mood, aspect, tense, and the persons of the verb's subject and object. But not all of these categories appear on all verbs. We have already seen several examples of simple verbs ending in [-n], the marker for a general present tense, not one specifically focused on right now – there is a separate present progressive. A verb ending in [-n] is understood to have a third person subject, and, if transitive, a third person object. Third person is the default for verbs and is not marked explicitly.

If the subject is some person other than third, a suffix is added. These are essentially the same markers than can appear as possessive personal suffixes on nouns. (54) shows the forms of [riku-] ‘see’ with subject suffixes:

- (54) a. [riku-n-i] see-PRS-1: ‘I see’ or ‘I see him’ or ‘I see her’ or ‘I see them’
 b. [riku-n-ki] see-PRS-2: ‘You see’ or ‘You see him’, etc.

Matters are a little more complicated when the verb has an object other than third person (55). You will notice that we don't have any cases of a subject and an object both having the same person; that requires a different construction, the middle, which we will discuss later.

- (55) a. 1>2: [riku-j-ki] 'I see you'
 b. 2>1: [riku-wa-n-ki] 'You see me'
 c. 3>1: [riku-wa-n] 'He sees me'
 d. 3>2: [riku-su-n-ki] 'He sees you'

In (55a), the tense marker [-n] is omitted. The others have a special morpheme inserted before the tense marker. This may be an indication that the expected agency hierarchy (1>2>3) has been inverted. [-wa] (1.INV) means that something lower in the hierarchy is acting on the first person, and [-su] (2.INV) means a third person is acting on the second person. Note that in 'he sees you' (55d), the normal second person suffix is also found after the tense marker. That does not happen for the first person object (55b–c).

A plural ending can be added to verbs. The endings are essentially the same as for the personal suffixes on nouns: usually [-ku], but [-tʃik] for forms including the second person, even as part of the first person inclusive ('you and I'). However, only one suffix can be used, even though there may be both a subject and an object to consider. Again, the agency hierarchy seems to come into play. If either the subject or object is first person exclusive, [-ku] is used. Otherwise, if either the subject or object is second person or first person inclusive, [-tʃik] is used. Otherwise, if all parties are third person, [-ku] is used. This keeps the verbs shorter than they might have been, at the cost of some confusion as to what exactly is plural. For example, [riku-j-ki-ku] means 'we see you', but there is no room for a marker to indicate whether the second person reference is singular or plural. Quechua speakers use standalone personal pronouns (Table 7.33) when they need to avoid such ambiguities.

The primary tenses of Quechua are marked by suffixes that are positioned before most of the personal suffixes:

- (56) a. [-n] simple present: [riku-n] 'she sees'
 b. [-nχa(n)] future: [riku-nχa] 'she will see'
 c. [-rχa(n)] simple past: [riku-rχa] or [riku-ra] 'she saw'
 d. [-sχa(n)] HEARSAY past: [riku-sχa] 'she seems to have seen'

The hearsay past is used to report a past event that the speaker cannot vouch for from personal experience. The parenthesized [n] appears before most personal suffixes: [riku-rχan-ki] 'you saw'.

The tense suffixes were obviously constructed from smaller parts – it can't be a coincidence that [χan] appears in three of them. Parker (1969) explains them as composed of an aspect marker plus a time marker. [χan], for example, would be a marker of non-present tense.

The habitual or iterative past is formed with a participial form in [-χ] plus an auxiliary verb 'be' in either the present or past tense. However, 'be' is dropped in the third person:

- (57) a. [riku-χ ka-ni] or [riku-χ ka-rχa-ni] 'I used to see'
 b. [riku-χ] 'he used to see'

In the future tense, only the third person has the form with [-nχa(n)]. In the first person, [-saχ] is used instead, without a separate person marker: [riku-saχ] 'I will see'. In the second person, the present tense forms fill in for the future.

The MIDDLE VOICE (MID) is formed by adding the suffix [-ku] to the verb stem. It indicates that the subject is affected by the verb. Sometimes this is the same as an English reflexive.

- (58) [riku-ku-n-i] see-MID-1SG 'I see myself', 'I see for myself'

It is sometimes also used as a passive. However, there is a native passive construction that is apparently not imported from Spanish and that does not use the middle voice:

- (59) paj=χa χawa-sχa ka-sχa huwansitu-wan=mi
 3SG=TOP see-NOUN be-NOUN Juancito-INS=EXPER
 'He was seen by Johnny.'

In addition to the indicative mood, which we have been discussing, Quechua has an imperative, a jussive, and a conditional. The imperative mood is signaled by the suffix [-j], which usually follows right after the stem (60a) or the inverse first person suffix (60b). The jussive forms third person commands.

- (60) a. [riku-j] 'See!'
 b. [janapa-wa-j] 'Help me!'
 c. [ri-tʃun] 'May he go!'

The conditional mood is signaled by the suffix [-man] added to the present tense of the verb, following all suffixes except the plural suffix [-ku] (61a). Besides the usual conditional meaning 'would', it may also mean 'should', 'may', and 'might'. The past conditional is made by adding the past tense of 'be' as an auxiliary verb (61b).

- (61) a. wakin-ku taki-n-man-ku
 other-PL sing-PRS[3]-COND-3PL
 'Others would sing.'
 b. tʃaj mesa-ta aʎin-ta ruwa-n-man ka-rχa
 that table-ACC good-ACC make-PRS[3]-COND be-PST
 'He should have made that table well.'

There are three different subordinate verb forms that function somewhat like English participles. The suffix [-pti] indicates that a verb is subordinated to another verb. The [p-] of this marker marks switch reference: the subject of this verb is different from the subject of the verb in the main clause. Therefore it must always be inflected for person. It indicates an action that begins before the action expressed by the verb in the main clause and is usually translatable into English by a clause introduced by ‘if’, ‘when’, or ‘because’.

- (62) ɲuxa ni-pti-j=mi $\text{ri-r}\chi\text{a}$
 1SG say-SUBORD.DS-1SG=EXPER go-PST
 ‘He went because I said it.’

The suffix [-spa] is different from [-pti] only in that its subject and the subject of the verb of the main clause are the same. Person marking on this form is optional.

- (63) miku-spa-n $\text{\textcircled{\scriptsize\Lambda}uxsi-r}\chi\text{a}$
 eat-SUBORD.SS-3SG leave-PST
 ‘Having eaten, he left’ or ‘When he had eaten, he left.’

The suffix [stin] is like [spa] but signals that its action takes place at the same time as the action of the main verb. This form is never inflected for person.

- (64) tuma-stin $\text{puriku-t}\overline{\text{\scriptsize\Lambda}}\text{kan-ki}$
 drink-SUBORD.SS walk_around-DUR.PRS-2SG
 ‘You are walking around drinking.’

Simultaneous action of two different actors must be expressed by a construction in which a [stin] subordinate form is in attributive relationship to a [pti] subordinate form of the durative stem of [ka-] ‘to be’:

- (65) waxa-stin $\text{ka-t}\overline{\text{\scriptsize\Lambda}}\text{ka-pti-n}$ $\text{\textcircled{\scriptsize\Lambda}uxsi-r}\chi\text{a-ni}$
 cry-SUBORD.SS be-DUR-SUBORD.DS-3 leave-PST-1[sg]
 ‘I left while she was crying.’

Verb derivation is very richly developed in Quechua. Verbs are derived from noun or adjective stems by means of various verbalizing suffixes:

- (66) a. [$\text{wasi-t}\overline{\text{\scriptsize\Lambda}}\text{a}$] house-v ‘to make a house’
 b. [$\text{suma}\chi\text{-ja}$] pretty-v ‘to become pretty’

According to Parker (1969: 31), Ayacucho Quechua has 16 common derivational suffixes that express grammatical features like aspect and voice. They can occur in combination with each other to form a large number of derived forms from a single verb stem. Only a few such derivations are listed here:

- (67) a. [riku-pa-n]: ‘he sees again’ (repetitive)
 b. [riku-raja-n]: ‘he sees continually’ (CONTINUATIVE)

- c. [riku-ru-n]: 'he has just seen', 'he sees urgently' (sudden or urgent action)
- d. [riku-tʃi-n]: 'he has someone see' (causative)
- e. [riku-ji-n]: 'he helps someone see' (auxiliary)
- f. [riku-pu-n]: 'he sees it for someone' (benefactive)
- g. [riku-na-ku-n]: 'he and someone else see each other' (reciprocal plus middle)
- h. [riku-mu-n]: 'he sees there', 'he goes to see' (location, motion)
- i. [riku-tʃka-n]: 'he is seeing' (durative or progressive)

Quechua does not have a verb meaning 'have'. Possession is indicated by using the verb 'be' with personal possessive suffixes on nouns or with the genitive case.

- (68) a. bisinti-pa=χa ka-tʃka-n-ɲa=s musuχ wasi-n
 Vicente-GEN=TOP be-DUR-PRS[3]-already=HSY new house-3SG
 'They say Vicente already has a new house.'
- b. lapis-nij ka-n
 pencil-1SG be-PRS[3]
 'I have a pencil.'

A more emphatic predication of possession is made by means of the benefactive verb suffix [-pu]:

- (69) lapis-nij ka-pu-wa-n
 pencil-1SG be-BEN-1SG.INV-PRS[3]
 'I have a pencil.'
 Literally, 'My pencil exists for me.'

7.27.3.5 Enclitics

There are some 13 enclitics in Quechua. They attach to the end of the word or phrase they apply to, and more than one can attach at the same time.

Typically, both the topic and the focus of sentences are marked explicitly. The topic [=χa] is what the discourse is about; the focus is the comment the speaker is making about the topic. The topic marker in statements is always [=χa] (70a; examples here are from Zariquiey & Córdova 2008). The focus markers vary depending on the confidence the speaker is asserting. [=m] means the speaker is asserting that she has first-hand information: she has personally experienced what she is talking about. [=s] is the opposite: the information is presumably true, but she heard it from somebody else, or experienced it when not clear-minded – perhaps when drunk or surprised, or when she was a small child. There is an obvious semantic connection with Quechua's hearsay past, which, you will recall, differs from the simple past in having an [s]. In our glosses, we summarize [=m] and [=s] as experiential (EXPER) versus hearsay (HSY) evidentiality markers. When they follow a consonant, they are extended to [=mi] and [=si].

- (70) a. [=χa] topic, [=m] experience:

karlus=χa jat[at]jix=mi
 Carlos=TOP professor=EXPER

‘Carlos is a professor’ (As for Carlos, I personally can assert as a fact that he is a professor.)

Note that even a verbless sentence has a predicate that can be in focus.

- b. [=s] hearsay:

karlus kunan tanta-ta miku-n=si
 Carlos today bread-ACC eat-PRS=HSY

‘They say that Carlos is eating bread today.’

The hearsay particle could be attached to any other word in the sentence to indicate that that is the part of the comment that is second-hand – that it is Carlos who is eating bread [karlus=si], that it is happening today [kunan=si], and so forth.

Here are some examples of other enclitics. In addition, the interrogative-negative enclitic [=tʃu] will be discussed under Syntax §7.27.4.

- (71) a. [=puni] ‘definitely’:

wasi=puni=m
 house=definitely=FOC

‘It is definitely a house.’

- b. [=ɲa] ‘already’:

karu=ɲa=m
 far=already=FOC

‘It is already far.’

This enclitic very often combines with [=taχ].

- c. [=taχ] sequential or contradictory:

jana=taχ
 black=contradiction

‘No, it is black!’

- d. [=ja] regret or resignation:

wasi-j=ja
 house-1SG=regret

‘Oh my poor house!’

- e. [-
- ¹
- a] emphasis:

wasi-n=¹s=a
 house-3SG=HSY=EMP

‘He says it is his house!’

When it does not follow a focus enclitic, this morpheme is realized simply as a shift of stress onto the last syllable of the word.

7.27.4 SYNTAX

7.27.4.1 Word order

The basic word order in Quechua is SOV, and modifiers usually precede the words they modify. However, adverbial phrases may often follow the verb they modify, and SVO order is not entirely uncommon, perhaps due to Spanish influence:

- (72) a. waɫpa-kuna sara-ta miku-tʃka-n-ku
 chicken-PL maize-ACC eat-DUR-PRS-3PL.SBJ
 SUBJECT OBJECT VERB
 ‘The chickens are eating corn.’
- b. antʃa aɫin wasi
 very good house
 Adv Adj N
 ‘a very good house’
- c. hawka jatʃa-n
 peaceful live-PRS[3]
 Adv VERB
 ‘He lives peacefully.’
- d. ri-n wajna-m-pa-ta
 go-PRS[3] love-3-GEN-ACC
 ‘She goes to her lover’s (house).’
 Adverbial phrase following the verb.

7.27.4.2 Grammatical relations

Grammatical relationships are signaled primarily by (73a) case suffixes, (73b) nouns with case suffixes used as postpositions, (73c) subject and object suffixes on verbs, and, more rarely, (73d) a few prepositions borrowed from Spanish. As in this last example, the Spanish prepositions are often redundant and optional.

- (73) a. wasi-pi
 house-LOC
 ‘in the house’
- b. wasi uku-pi
 house interior-LOC
 ‘inside the house’
- c. riku-j-ki
 see-1-2
 ‘I see you.’
- d. asta uktubri-kama
 until October-TERM
 ‘until October’

7.27.4.3 Agreement

Gender agreement is foreign to Quechua, but a small number of nouns borrowed from Spanish preserve sex-based gender distinctions, and an even smaller number of borrowed adjectives actually show gender agreement:

- (74) a. [amigu] ‘friend’ M vs. [amiga] ‘friend’ F
 b. [loko] ‘crazy’ (with a masculine noun) vs. [loka] ‘crazy’ (with a feminine noun)

Agreement in person is mandatory between the subject of a verb and the subject suffix attached to the verb:

- (75) χ am χ a atʃka-ta=m miku-n-ki
 2SG much-ACC=EXPER eat-PRS-2SG
 ‘You eat a lot.’

However, the subject and the verb do not have to agree in number in all situations. Thus both of these sentences are acceptable:

- (76) ‘The men are playing.’
 a. runa-kuna puk λ a-tʃka-n-ku
 person-PL play-DUR-PRS-PL
 b. runa-kuna puk λ a-tʃka-n
 person-PL play-DUR-PRS

*[runa puk λ a-tʃka-n-ku] is not grammatical, because it is not possible to have a plural subject suffix on the verb when its subject is not overtly marked for plurality.

7.27.4.4 Subordinate clauses

Instead of being introduced by various subordinate conjunctions as in English, Quechua subordinate clauses for the most part are based on participle-like subordinate verb forms:

- (77) miku-spa-n λ uxsi-r χ a
 eat-SUBORD.SS-3 leave-PST[3]
 ‘When he had eaten he left.’

For similar subordinate verb constructions, see again §7.27.3.4.

Other types of subordinate clauses involve various nominalized verb forms:

- (78) jatʃa-n-i wasi ruwa-na-n-ta
 know-PRS-1SG house make-NOUN.POT-PRS-ACC
 ‘I know that he will build a house.’

7.27.4.5 Relative clauses

Relative clauses are not introduced by relative pronouns, unlike in English and Spanish. Quechua equivalents of relative clauses are built around a verb that has been nominalized by one of the Quechua nominalizing suffixes, some of which have specialized functions. For example, [-χ], a suffix that derives agent or instrument nouns from verbs, can be used in the following construction to render one type of a relative clause:

- (79) *Λamka-χ runa*
 work-NOUN.AGT person
 ‘a man who works’

Many other noun- and adjective-building suffixes create words that may be translated into English as relative clauses:

- (80) a. [wasi-juχ] house-possessor ‘person who owns a house or houses’, ‘a landlord’
 b. [wasi-sapa] house-many_possessor ‘a man who has many houses’
 c. [wasi-ntin] house-adjacent ‘that which has a house next to it’

7.27.4.6 Interrogative sentences

Polar questions are formed by adding the interrogative enclitic [=t̃ju] to the word being questioned (like [l̃i] in literary Russian) and moving the questioned word to the front of the sentence:

- (81) a. *ajakut̃ju-ta=t̃ju paχarin ri-nχa-ku*
 Ayacucho-ACC=Q tomorrow go-FUT-PL
 ‘Will they be going to Ayacucho (and not some other place) tomorrow?’
 b. *paχarin=t̃ju ajakut̃ju-ta ri-nχa-ku*
 tomorrow=Q Ayacucho-ACC go-FUT-PL
 ‘Is it tomorrow that they will be going to Ayacucho?’

Negative polar questions add the [=t̃ju] enclitic to the negative morpheme:

- (82) *mana=t̃ju hamu-n-ki*
 NEG=Q come-PRS-2SG
 ‘Aren’t you coming?’

Content questions use interrogative words and usually do not add the interrogative enclitic:

- (83) *ima-ta=taχ tajta-jki ruwa-t̃jka-n*
 what-ACC=FOC father-2SG do-DUR-PRS
 ‘What is your father doing?’
 [=taχ] is a focus marker in questions.

7.27.4.7 Negation

Negation is signaled by the negative morpheme [mana] or the prohibitive marker [ama]. In addition, the negative enclitic [=tʃu] is added to the main verb or to the noun or pronoun that acts as a predicate in a verbless sentence:

- (84) a. mana=m ri-nχa=tʃu
 NEG=EXPER go-FUT=NEG
 ‘He won’t go.’
- b. mana ri-spa-n
 NEG go-SUBORD.SS-3
 ‘not going, without going’
- c. mana=m wasi=tʃu
 NEG=EXPER house=NEG
 ‘It is not a house.’
- d. mana wasi-juχ
 NEG house-POSS
 ‘one who has no house, houseless’

Negative words such as ‘nobody’ and ‘nothing’ are formed with interrogative words acting as indefinite pronouns. These interrogatives are negated and then followed by an enclitic meaning ‘even’. This construction is very similar to its Japanese equivalent.

- (85) mana pi=pas rima-n=tʃu
 NEG who=even speak-PRS=NEG
 ‘Nobody speaks.’
 Literally, ‘Not anyone speaks.’

7.27.5 SAMPLE TEXT

This sample text is from Parker (1969), who did not indicate where he got it from. The Quechua text as well as its idiomatic translation, both of which are quoted here, are to be found at the end of Parker’s grammar. It is published here with the permission of Mouton de Gruyter. The morpheme-by-morpheme annotation and the accompanying notes are our own.

- (86) Adriyan Warmacha
 adrijan warma-tʃa
 Adrian boy-DIM
 ‘The Little Boy Adrián’

- (87) Huk siñuras huk warmata uywasqa wasimpi yanapanampaq, Adriyan sutiuyuqta.
 huk siñur-a=s huk warmata-ujwa=sχa wasi-m-pi
 a_certain lady-F=HSY a_certain boy-ACC raise=HSY.PST house-3SG-LOC
 janapa-na-m-paχ adrijan suti-juχ-ta
 help-NOUN-3SG-BEN Adrian name-possessor-ACC

‘Once a woman kept a boy named Adrián in her house to help her.’

[wasi-m-pi]. Before bilabial consonants, the [-n] of various suffixes regularly becomes [-m] by assimilation.

[adrijan suti-juχ-ta]. The entire noun phrase is in apposition to [warmata] ‘boy’: ‘boy . . . Adrian name possessor’ = ‘boy . . . Adrian by name.’

- (88) Siñuraqa kamachiq llapa imata chay warmachata,
 siñur-a=χa kamat̃ji-χ lapa ima-ta t̃jaj warmat̃ja-ta
 lady-F=TOP command-ITER.PST all what-ACC that(DEM) boy-DIM-ACC
 ‘The woman ordered the little boy to do everything.’

- (89) warmachañataq mana imatapas ruwaqchu allintaq.
 warmat̃ja=jataχ mana ima-ta=pas ruwa-χ=t̃ju aλin-ta=χa
 boy-DIM=then NEG what-ACC=even do-ITER.PST=NEG good-ACC=TOP
 ‘but the boy never did anything well.’

- (90) Sapa triguta akllaspan, kutaspan, wakiwakillanta akllaykuq, kutaykuq;
 sapa trigu-ta akλa-spa-n kuta-spa-n waki-waki-λan-ta
 every wheat-ACC choose-SUBORD.SS-3 grind-SUBORD.SS-3 bit-bit-only-ACC
 akλa-jku-χ kuta-jku-χ
 choose-oddly-ITER.PST grind-oddly-ITER.PST

‘Every time he picked out some wheat and ground it, he picked it out and ground it just a little at a time.’

[akλa-jku-χ]. It is usually not possible to translate [-jku] into English. In many cases the meaning is very idiomatic: [rima-] ‘to speak’, but [rima-jyku-] ‘to greet’. According to Parker (1969: 67), it may indicate cordiality, severity, fear, surprise, and so on, depending on context.

- (91) wakintañataq maraypa ladun quchaman wischuykuq.
 waki-n-ta=jataχ maraj-pa ladun χut̃ja-man wist̃ju-jku-χ
 bit-3SG-ACC=then millstone-GEN side well-ALL throw_away-oddly-ITER.PST
 ‘and he threw away a part of it into the well beside the millstone.’

- (92) Mikuykunatapas wischuykariq wakillanta mikuspan.

mikuj-kuna-ta=pas wistʃu-jka-ri-χ waki-ʎan-ta miku-spa-n
 food-PL-ACC=also throw_away-oddly-INCHO-ITER.PST bit-only-ACC eat-SUBORD.SS-3

‘He also threw away food after eating only part of it.’

[wistʃu-jka-ri-χ]. The form [-jka-] is the allomorph of [-jku-] before [-mu] and a few other suffixes.

- (93) Warmaqa pukllayla pukllakuq sapa punchaw mana kasukuspan.

warma=χa pukʎay-la pukʎa-ku-χ sapa puntʃaw mana kasu-ku-spa-n
 boy=TOP play-only play-MID-ITER.PST every day NEG obey-MID-SUBORD.SS-3

‘The boy just played and played every day without ever being obedient.’

- (94) Qullqitawan pukllanakunata suwakamun bisinun wasikunamanta,

χuʎi-ta-wan pukʎa-na-kuna-ta suwa-ka-mu-n bisinu-n wasi-kuna-manta
 money-ACC-INS play-NOUN-PL-ACC steal-MID-move-PRS[3] neighbor-3SG house-PL-ABL

‘He would go steal money and toys from his neighbors’ houses.’

[suwa-ka-mun]. The form [-ka-] is the allomorph of [-ku-] before [-mu] and a few other suffixes.

- (95) hinaspan maray qipapi pakan sapa patrunan llusiptin pukllanampaq.

hina-spa-n maraj χipa-pi paka-n sapa patrun-a-n
 be_thus-SUBORD.SS-3 millstone behind-LOC hide-PRS[3] each boss-F-3SG

ʎuxsi-pti-n pukʎa-na-m-paχ
 depart-SUBORD.DS-3 play-NOUN-3SG-BEN

‘and then hide them behind the millstone, to play with each time his boss left.’

[paka-n]. In narratives, or indeed any time that the tense is obvious, the present tense is often used.

- (96) Kayna kachkaspan, warmachaqa unqurun yana mumwan;

kaj-na ka-tʃka-spa-n warma-tʃa=χa unχu-ru-n jana mum-wan
 this-NOUN be-DUR-SUBORD.SS-3 boy-DIM=TOP get_sick-sudden-PRS black seed-INS

‘In this state of affairs the boy got sick with smallpox.’

- (97) hinaspan wañukun chay unquywan manaña imawampas allinyayta atispan.

hina-spa-n waju-ku-n tʃaj unχu-j-wan mana=ña ima-wam=pas
 be_thus-SUBORD.SS-3 die-MID-PRS that get_sick-NOUN-INS NEG=more what-INS=even

aʎin-ja-j-ta ati-spa-n
 good-become-NOUN-ACC be_able-SUBORD.SS-3

‘and then he died of this disease before he could reform at all.’

- (98) chay wañusqanmanta pacha, manchachikuyta qallaykun huk yana bultu qucha chawpipi sapa tuta.

tʃaj waju-sʃa-n-manta patʃa mantʃatʃi-ku-j-ta ʃaʃa-jku-n
that die-PST.NOUN-3SG-ABL time frighten-MID-NOUN-ACC begin-oddly-PRS

huk jana bultu ʃutʃa tʃawpi-pi sapa tuta
a black spirit well middle-LOC each night

‘Soon after he died, a black spirit began to frighten (people) around the well each night.’
[wajusʃanmanta patʃa]. Literally, ‘time from his having died’ = ‘since his death.’

- (99) Kayta rikuruspa warmiqa mancharikun, hinaspan rin tayta kuraman willaq.

kaj-ta riku-ru-spa warmi=ʃa mantʃari-ku-n
this-ACC see-suddenly-SUBORD.SS woman=TOP be_afraid-MID-PRS

hina-spa-n ri-n tajta kura-man wiʃa-ʃ
be_thus-SUBORD.SS-3 go-PRS[3] father priest-ALL tell-PTCP

‘Seeing this, the woman became frightened, and she went to the priest to tell him.’

- (100) Tayta kurañataq yachachin krusta hapiykuspa pim kasqanta tapunampaq.

tajta kura-ɲataʃ jatʃatʃi-n krus-ta hapi-jku-spa pi=m
father priest-then teach-PRS cross-ACC hold-only-SUBORD.SS who=EXPER

kasʃan-ta tapu-na-m-paʃ
again-ACC ask-NOUN-3SG-BEN

‘The priest taught her to hold a cross and ask who it was.’

- (101) Warmiqa kutiykun wasinta chay ruwaq,

warmi=ʃa kuti-jku-n wasi-n-ta tʃaj ruwa-ʃ
woman=TOP return-oddly-PRS house-3SG-ACC this do-PTCP

‘The woman returned home to do that,’

- (102) tutachaykuptin suyan chay yana bultu rikurimunanta.

tuta-tʃa-jku-pti-n suja-n tʃaj jana bultu
dark-v-oddly-SUBORD.DS-3 wait_for-PRS[3] this black spirit

riku-ri-mu-na-n-ta
see-INCHO-move-NOUN-3SG-ACC

‘and when it became dark she waited for the black spirit to appear.’

- (103) Rikuriramuptin, mancharikuspan pampaman kumpakun;

riku-ri-ra-mu-pti-n mantʃa-ri-ku-spa-n
see-INCHO-suddenly-move-SUBORD.DS-3 frighten-INCHO-MID-SUBORD.SS-3

pampa-man kumpa-ku-n
ground-ALL throw_down-MID-PRS[3]

‘When it appeared she became frightened and threw herself to the ground,’

- (104) *hinaspa chayna pampapi wischusqa achikyarun.*
 hina-spa t̃ajna pampa-pi wist̃ju-s̃xa at̃jikja-ru-n
 be_thus-SUBORD.SS do_thus ground-LOC abandon-PST.PTCP dawn-suddenly-PRS[3]
 ‘and she stayed like that lying on the ground till dawn.’
- (105) *Paqarinnintinta kaqlla rin tayta kurapata, hinaspa kikinta pusaramun.*
 paɣari-nnintin-ta kaɣʌa ri-n tajta kura-pa-ta hina-spa
 daybreak-next-ACC again go-PRS[3] father priest-GEN-ACC do_thus-SUBORD.SS
 kiki-n-ta pusa-ra-mu-n
 self-3SG-ACC lead-urgently-motion-PRS[3]
 ‘On the following day she went again to the priest’s, and brought him himself.’
- (106) *Kuraqa krusta apaykuspa suyan,*
 kura=ɣa krus-ta apa-jku-spa suja-n
 priest=TOP cross-ACC take-oddly-SUBORD.SS await-PRS
 ‘The priest waited carrying a cross.’
- (107) *rikuriramuptinñataq tapun “Pim kanki, imatam munanki?” nispan.*
 riku-ri-ra-mu-pti-n=ɲataɣ tapu-n pi=m ka-n-ki
 see-INCHO-urgent-move-SUBORD.DS-3=then ask-PRS[3] who=EXPER be-PRS-2SG
 ima-ta=m muna-n-ki ni-spa-n
 what-ACC=EXPER want-PRS-2SG say-SUBORD.SS-3
 ‘and when it appeared he asked: “Who are you, and what do you want?”’
- (108) *Chay bultu kontestan: “Adriyanmi kani, taytacham kutichimuwan;*
 t̃aj bultu kontesta-n adrijan=mi ka-n-i tajta-t̃ja=m
 that spirit answer-PRS Adrian=EXPER be-PRS-1SG father-DIM=EXPER
 kutit̃ji-mu-wa-n
 send_back-move-1SG.INV-PRS
 ‘The spirit answered: “I am Adrián; God has sent me back here;’
- (109) *kay quchapim llapa ima tallisqay kachkan,*
 kaj ɣut̃ja-pi=m ʌapa ima taʌi-s̃xa-j ka-t̃jka-n
 this well-LOC=EXPER all what throw_out-PST.PTCP-1[sg] be-DUR-PRS
 ‘in this well is everything I threw away’
- (110) *wak maray qipapiñataqmi llapa suwakusqa pakasqay.*
 wak maraj ɣipa-pi=ɲataɣ=mi ʌapa suwa-ku-s̃xa paka-s̃xa-j
 yonder millstone behind-LOC=then=EXPER all steal-MID-PST.PTCP hide-PST-1SG
 ‘and behind that millstone I hid everything I stole.’

- (111) Ama hina kaspaykichik urquruychik, mana chayqa supay wasipaqmi taytacha unanchawanman.”
 ama hina ka-spa-j-ki-tʃik urχu-ru-j-tʃik mana tʃaj=χa supaj
 PROH such be-SUBORD.SS-1SG-2-PL take_out-urgently-IMP-PL NEG that=TOP devil
 wasi-paχ=mi tajta-tʃa unantʃa-wan-man
 house-BEN=EXPER father-DIM destine-1.INV-COND
 “Please take it out, otherwise God would destine me for hell.”
 [ama hina ka-spa-j-ki-tʃik]. An idiom for ‘please.’
- (112) Chay rratulla urquyta qallaykunku, ña tukuruptinkuñaataq tayta kura nin:
 tʃaj zatuʎa urχu-j-ta χaʎa-jku-n-ku ja
 that immediately take_out-INF-ACC begin-oddly-PRS-PL already
 tuku-ru-pti-n-ku=jataχ tajta kura ni-n
 finish-urgently-SUBORD.DS-3-PL=then father priest say-PRS
 ‘Immediately they began to take it out, and when they had finished the priest said:’
- (113) “Kananqa hawka kutiriy taytachaman, ñam urqurunikuña.”
 kanan=χa hawka kuti-ri-j tajta-tʃa-man ja=m
 now=TOP peaceful return-INCHO-IMP father-DIM-ALL already=EXPER
 urχu-ru-ni-ku=ja
 take_out-urgently-1-PL.EXCL=already
 “Now return in peace to God; we have taken it out.”
- (114) Chaynapi chay yana bultuqa chinkarikurun “grasyas” nispan.
 tʃaj-na-pi tʃaj jana bultu=χa tʃinka-ri-ku-ru-n grasjas
 that-NOUN-LOC that black spirit=TOP disappear-INCHO-MID-suddenly-PRS thanks
 ni-spa-n
 say-SUBORD.SS-3
 ‘Then the black spirit disappeared, saying “Thank you.”’

7.28 Exercises

7.28.1 HIXKARYÁNA AND LANGUAGE TYPOLOGY

Hixkaryána §7.25 is spoken by a small group of people in the Amazonas province of Brazil, not far from Manaus. Its basic word order, OVS, is rare among the languages of the world, and at one time many linguists believed that it did not occur at all. For that reason this language is of great interest to language typologists.

Carefully examine the Hixkaryána data given below (from Derbyshire 1979 and 1985) and then try to answer the following questions.

- a. How are grammatical relations signaled in Hixkaryána? In your answer be sure to indicate how you reached your conclusions.
- b. What is the order of the various types of modifiers in this language? In what ways is this language's order of constituents similar to that of a typical SOV language, and in what ways is it different?

- (1) namtjeh^htjowuu totokomo
3.hunt.PL.PST person.PL
'The people have gone hunting.'
- (2) totokomo jonet^hkonuu roro hatuu kamar-jana komo
person.PL 3>3.eat.PL.PST permanent HSY jaguar-person PL
'The jaguar people used to eat people all the time.'
- (3) asak kanawa wejo
two canoe 1>3.see.PST
'I saw two canoes.'
- (4) anaro owto hona kahatakeko
another village to (1).come_out.PST
'I arrived at another village.'
- (5) burjekomo joknuu
boy pet.POSSESSED
'the boy's pet'
- (6) burjekomo jotahano ujo
boy 3>3.hit.PST her_husband
'Her husband hit the boy.'
- (7) toto jahosuje kamara
man 3>3.grab.PST Jaguar
'The jaguar grabbed the man.'
- (8) toto jahosuje
man 3>3.grab.PST
'It grabbed the man.'
- (9) kamara nahosuje
jaguar 3>3.grab.PST
'The jaguar grabbed him.'
The order [nahosuje kamara] is also possible.
- (10) toto nahosuje
man 3>3.grab.PST
'The man grabbed it (the jaguar).'
The order [nahosuje toto] is also possible.

- (11) nahosujje
 3>3.grab.PST
 'It grabbed him.' or 'He grabbed it.'
- (12) ehonomnu me rmahaja naha jaskomo totokomo wja
 important_one DENOMINAL very_much 3.is shaman people to
 'The people think the shaman is a very important person.'
- (13) uwahathujamo akrohpenhujamo tho oske nketʃkonu
 his_killers ones_who_had_burned_him DEVALUED thus 3.say.PL.PST
 'His killers, the ones who had burned him, said thus.'
- The DEVALUED particle marks the items it modifies as having undergone some change of state or relationship, usually involving loss of value.
- (14) ʃetʃa waha ntetʃhe onokna komo
 forest through 3.go.PL creature PL
 'Creatures go through the forest.'

7.28.2 MICHOACÁN NAHUATL MORPHOLOGY

Perform a morphological analysis on these sentences of Michoacán Nahuatl (NCL). For each morpheme, say whether it is a root, prefix, or suffix, and identify its function. (This exercise is Problem 179 in Merrifield et al. 1987. Used with permission.)

- | | |
|----------------------------|--------------------------|
| 1. nikotʃik | I slept. |
| 2. kotʃik iʃolul | His child slept. |
| 3. tijuli | You live. |
| 4. juli mosiwal | Your wife lives. |
| 5. netʃlamatʃiltia | He informs me. |
| 6. kilamatʃiltik nosiwal | He informed my wife. |
| 7. tiwehkawa | You endure. |
| 8. wehkawa notʃkawalisli | My strength holds out. |
| 9. nilamik | I finished. |
| 10. lami molamatʃiltilisli | Your news ends. |
| 11. lamik itʃikawalisli | His strength gave out. |
| 12. mihtʃikawak | He strengthened you. |
| 13. kitʃikawa | He strengthens him. |
| 14. kitʃikawa noʃolul | He strengthens my child. |
| 15. kipolua | He loses it. |
| 16. kipolua kotʃililisi | He loses sleep. |
| 17. kipoluk itʃikawalisli | He lost his strength. |
| 18. netʃwililtia | He empowers me. |

- | | | |
|-----|-----------------------|-------------------------|
| 19. | mitʃwilitik | He empowered you. |
| 20. | kiwilitia moʃolul | He empowers your child. |
| 21. | netʃneki | He loves me. |
| 22. | kineki julilisli | He wants life. |
| 23. | kineki isiwal | He loves his wife. |
| 24. | kinetik nowililitisli | He wanted my power. |

7.28.3 HUAVE MORPHOLOGY AND SYNTAX

Examine the Huave data given below and then answer the following questions. For the sake of simplicity, tones should be ignored and are therefore not marked. This exercise is Problem 250 in Merrifield et al. (1987: 178). Used with permission.

- What is the basic word order of this language?
- How does it signal grammatical relations?
- What is the morphological structure of the Huave verb?
- To what general language type does it belong? Be sure to cite specific examples in support of your claim.

- | | | |
|-----|--|--------------------------------|
| 1. | ta ⁿ dok tiʃem | He netted shrimp. |
| 2. | taho ⁿ d ki ^e t nop naʃej | A man dried fish. |
| 3. | teha ⁿ ts | He washed (something). |
| 4. | apmaha ⁿ ts ki ^e t | He will wash fish. |
| 5. | a:ga ne ⁿ tʃ apmeha ⁿ ts | That boy will wash something. |
| 6. | apma ⁿ dok tiʃem a:ga ne ⁿ tʃ | That boy will net shrimp. |
| 7. | nop ne ⁿ tʃ te ⁿ dok | A boy netted (something). |
| 8. | a:ga naʃej teho ⁿ d | That man dried (something). |
| 9. | nop ne ⁿ tʃ apmaho ⁿ d ki ^e t | A boy will dry fish. |
| 10. | nop naʃej apmeho ⁿ d | A man will dry (something). |
| 11. | a:ga naʃej taha ⁿ ts tiʃem | That man washed shrimp. |
| 12. | apme ⁿ dok | He will net (something). |
| 13. | nop naʃej apma ⁿ dok ki ^e t | A man will net fish. |
| 14. | taho ⁿ d tiʃem | He dried shrimp. |
| 15. | nop naʃej taho ⁿ d ki ^e t | A man dried fish. |
| 16. | a:ga ne ⁿ tʃ apmeho ⁿ d | That boy will dry (something). |
| 17. | nop ne ⁿ tʃ apmeha ⁿ ts | A boy will wash (something). |
| 18. | a:ga ne ⁿ tʃ apma ⁿ dok tiʃem | That boy will net shrimp. |

7.29 Suggested readings

7.29.1 GENERAL

- ❏ *American Indian languages: The historical linguistics of native America* (L. Campbell 1997). Comprehensive scholarly account of the origins and historical development of the languages of the Americas.

- ❏ *Amerindian syntax* (E. Cook & Gerdtts 1984).
- ❏ *Language in the Americas* (Greenberg 1987). Defends the controversial “Amerind” hypothesis, that there is a clade that includes all Native American Indian languages except Na-Dene and Eskimo-Aleut.
- ❏ Review of *Language in the Americas* by Joseph A. Greenberg (L. Campbell 1988). Campbell’s review illustrates why Greenberg’s 1987 classification has not gained wide acceptance, due to what is seen as loose application of questionable methods.

7.29.2 NORTH AMERICA

- ❏ *The languages of native North America* (Mithun 1999). Comprehensive survey of languages, including their classification and their diverse typological characteristics.
- ❏ Native American languages (Yamamoto & Zepeda 2004). Introductory-level overview of North American indigenous languages and the main linguistic features that distinguish them from European languages.
- ❏ Overview of general characteristics (Mithun 1996). Concise overview of indigenous languages of North America, with special attention to linguistic characteristics that differentiate them from other languages.
- ❏ Survey of California and other Indian languages (<http://linguistics.berkeley.edu/~survey>). Primary materials, field notes, and linguistic descriptions of 130 languages, including 50 languages of California that are either severely endangered or extinct.

7.29.3 ALASKA

- ❏ Alaska Native Language Center (<http://www.uaf.edu/anlc/>). Digital library of materials on the 20 native languages of Alaska. Included are collections of stories in the languages, grammars, dictionaries, and linguistic descriptions.
- ❏ *Alaska native languages: Past, present, and future* (Krauss 1980).
- ❏ *A practical grammar of the Central Alaskan Yup’ik Eskimo language* (Jacobson & Jacobson 1995). A 548-page grammatical descripton in the form of a traditional-style language textbook.

7.29.4 MESOAMERICA

- ❏ Middle American languages (L. Campbell 1979).
- ❏ The archive of the indigenous languages of Latin America (<http://www.ailla.utexas.org>). A vast digital archive of oral and written source material from indigenous languages of Central and South America, plus some scholarly works based on the materials.

7.29.5 SOUTH AMERICA

- ✘ *The Amazonian languages* (Dixon & Aikhenvald 1999). Introduction covering the cultures of speakers, areal features of the languages, and genetic relationships among them, followed by chapters describing the grammar of individual language families.
- ✘ *Handbook of Amazonian languages* (Derbyshire & Pullum 1986).
- ✘ Language history in South America: What we know and how to know more (Kaufman 1990). Survey of the state of knowledge of language history and language classification in South America.
- ✘ *Tarma Quechua: Grammar, texts, dictionary* (Adelaar 1977). Reference grammar plus a dictionary, over 500 pages in all.

Language Birth, Death, and Revitalization

We have now completed our survey of the language families of the world. This cladistic approach works very well for the vast majority of languages, but there remain hundreds whose history is not well elucidated by the cladistic model. In this chapter we turn to those languages and study the special circumstances that affect their life cycle.

The basic cladistic model of language evolution assumes that a language has a base of speakers who use it as their general means of communication within their community. Their children acquire the language from them and other members of the community in a natural fashion and go on to be fluent native speakers of the language. Groups of speakers – typically, younger generations – will constantly change the language, but the changes will tend not to be so great or so abrupt as to impede communication with others. If different changes are systematically accepted in different areas, the language may eventually split into different languages. When such a split happens, the old language “dies” and two or more new ones are “born”.

In this chapter we look at other situations – languages that fall outside the standard narrative by coming into being without a split of the usual sort or that die without splitting into new languages.

8.1 Mixed languages

It is normal for languages in contact to influence one another. The most common way is through borrowing words. You are probably used to cases in which languages borrow words for new scientific and technological concepts, but languages also borrow ordinary vocabulary. Tadmor et al. (2012: 40) reported that 24% of the words they encountered in a large-scale study of 41 languages were loanwords. Native speakers are often unaware of which words are loans. In a short English-language article we selected blindly, which turned out to be a Céline ad about Kim

Kardashian walking down the street, we counted 101 different words, of which 38 were loanwords. Here's a sample from that ad, italicizing the loans:

- (1) *We just imagine how fabulous that fur feels when protecting one from those bitterly cold gusts of winter wind.*

Syntactic structures, too, can be altered through language contact. Many European languages characteristically draw their relative pronouns, including English *who* and *which*, from question words. This syntactic feature, not widespread in the world's language families, can spread to other languages through contact with languages from Europe. Nowadays this is happening in a variety of cases, including Tariana (TAE), an Arawakan language of Brazil in contact with Portuguese. In this language, younger speakers are now using the question word *kwana* 'who' to introduce relative clauses (Heine & Kuteva 2006: 213). Despite all this borrowing, languages retain their original lineage and remain in their original clade. There is no doubt that English remains a Germanic language and Tariana remains Arawakan.

In a few extreme cases, such influences are so profound that a distinct new language comes into being. MIXED LANGUAGES fuse materials from two languages so thoroughly that it is meaningless to classify them as having a single parent, per the normal cladistic model: they have two parent languages. This happens only when a community is thoroughly bilingual. Michif, which is spoken in North Dakota and nearby parts of Canada, is a combination of French and Plains Cree. With some exceptions, the two languages contributed distinct types of material to Michif. Most nouns, adjectives, and determiners are from French, with French-style gender agreement, masculine versus feminine; but demonstratives ('this', 'that') are from Cree and show Cree-style agreement: animate versus inanimate. Verbs are overwhelmingly of Cree origin and retain almost the full polysynthetic complexity of that language's verb morphology. Another mixed language is Media Lengua, which is spoken in Ecuador. Most of its lexical stems are of Spanish origin, but they take virtually the full panoply of suffixes found in Northern Quechua. As we saw in the Sketch of Ayacucho Quechua (§7.27), Quechua proper has borrowed many words from Spanish, but the situation in Media Lengua, which has dropped most of its native lexical roots, goes far beyond that. Here is a typical sample from Media Lengua, where all the word stems are from Spanish and all the suffixes are from Quechua:

- (2) Media Lengua

todabia no bjen aprendi-naku-n porke eskwela-bi anda-naku-n
 still not well learn-PL-3 because school-LOC go-PL-3
 'They don't learn well yet because they go to school.' (Muysken 1997: 401)

Table 8.1 lists a few mixed languages, along with their source languages.

TABLE 8.1
Some mixed languages

Name	Main sources	Location
Mednyj Aleut (MUD)	Aleut, Russian	Russia
Mbugu (MHD)	Bantu, Cushitic	Tanzania
Light Warlpiri	Warlpiri, English	Australia
Michif (CRG)	Plains Cree (CRK), French	Canada
Media Lengua (MUE)	Quechua, Spanish	Ecuador

8.2 Constructed languages

Languages can be born by being constructed intentionally, usually by one individual. Chances are you have heard of Esperanto (EPO), which was created to serve as a universal language that would be easy for everyone to learn and use. Devised in the late nineteenth century, Esperanto still has a serious following and is the most successful language ever constructed. According to some estimates, two million people use it as a second language. Classes are offered in various parts of the world. There is even an active Esperanto Wikipedia, <http://eo.wikipedia.org>.

Esperanto was one of hundreds of languages invented to foster communication among speakers of different languages in different countries. Languages created for such purposes are called *constructed international auxiliary languages*. Others have been created for philosophical or artistic purposes. Lojban (JBO), for example, was created in the 1990s to explore the idea of a language that would be syntactically unambiguous and based on predicate logic. Klingon (TLH) was invented to make the Star Trek movies more realistic than the original television series, in which the Klingon space aliens all spoke English. Its inventor, the linguist Marc Okrand, intentionally incorporated several properties that are typologically rare among Earth languages, such as OVS word order (recall our discussion of Hixkaryána in 7.25). To highlight the Klingons' hypermasculine warrior culture, the vocabulary of this language emphasizes war terminology, and its phonology is rich in harsh sounds produced in the velar and uvular region, such as [q̠χ]. At about the same time, in the 1980s, the linguist and novelist Suzette Haden Elgin invented Láadan (LDN) as an exercise in feminist language construction. The vocabulary of this language concentrates on concepts that the inventor felt were more important to women. Láadan is VSO and is completely devoid of velar and uvular consonants.

Constructed languages fall into a special category not because of their intrinsic properties but because of their origin: they did not arise from the usual process

of gradual change as language was passed down from generation to generation, but were engineered at an abrupt moment in time. The languages themselves vary widely, with some seeming very much like natural languages. If we didn't know Interlingua (INA) was developed in the 1940s by the International Auxiliary Language Association, we might well have thought it was a natural Romance language (3, with our own translation). Thus, linguistic properties alone may not definitively identify a language as constructed.

(3) Interlingua

- a. Nove inventiones technic crea nove parolas, anque in interlingua, que seque le developpamento natural del linguas que livra le vocabulario a interlingua, a saper anglese, germano, francese, espaniol, portugese, italiano e russo. Un aparato multo popular pote sonar musica conservate in forma digital in un formato special que le computadores pote leger, i.e. MP₃ (pronunciate m-p-tres). (Breinstrup 2013)
- b. New technological inventions also create new words in Interlingua, which follows the natural development of the languages that supply Interlingua with its vocabulary, namely English, German, French, Spanish, Portuguese, Italian, and Russian. A very popular device can play music digitally stored in a special format that computers can read, i.e. MP₃ (pronounced [empetres] [in Interlingua]).

Occasionally, languages encountered in the real world, so to speak, have strong signs of being constructed. The Lardil and Yangkaal tribes in Australia used to have a language called *Damin*. This language was taught to young men who underwent the subincision ritual, and thus was not learned as a native language by children. It had a much smaller vocabulary than the ordinary Lardil (LBZ) and Yangkaal languages and very different phonology. It had a bilabial trill [B], a velar ejective [k'], an ingressive lateral fricative with egressive glottalic release (maybe [ɬ↓ʔ]), and a series of clicks, including an egressive bilabial click [⊙↑] (Hale & Nash 1997). Many of its sounds are not found elsewhere in Australia, clicks are otherwise unique to Africa, and [ɬ↓ʔ] may well be found in no other language. It is hard to imagine how such a language could have arisen unless it was consciously invented by Lardil or Yangkaal initiates with a view toward sounding awesome.

8.3 Pidgins

An American merchant and student of languages, William C. Hunter, presented the following as representative of his conversations about taxation with the extremely wealthy Chinese businessman Houqua in Guǎngzhōu, China, in the 1830s (Hunter 1938: 22):

(4) Chinese Pidgin English

- a. Hunter: Well, Houqua, have got news today?
- b. Houqua: Have got too muchee bad news. Hwang Ho have spillum too muchee.
- c. Hunter: Mantale have come see you?
- d. Houqua: He no come see my, he sendee come one piece chop. He come tomollo. He wanchee my two lac dollar.
- e. Hunter: You pay he how muchee?
- f. Houqua: My pay he fitty, sikky tousand so.
- g. Hunter: But spose he no contentee?
- h. Houqua: Spose he, number one, no contentee, my pay he one lac.

(5) Our translation:

- a. Hunter: Well, Houqua, do you have any news today?
- b. Houqua: I have a surfeit of bad news. The Yellow River has flooded.
- c. Hunter: Has the mandarin (government official) come to see you?
- d. Houqua: He hasn't come to see me – he sent me a letter. He's coming tomorrow. He wants two hundred thousand dollars from me.
- e. Hunter: How much will you give him?
- f. Houqua: I'll give him fifty, sixty thousand or so.
- g. Hunter: But what if he's not satisfied?
- h. Houqua: If he, the boss, isn't satisfied, I'll give him one hundred thousand.

This language was called *Pigeon English*; in present terminology, it is Chinese Pidgin English (CPE). It originated in the years before the First Opium War, when Europeans were allowed to work and live only in very circumscribed areas near ports, principally Guǎngzhōu. Few Chinese at the time had interest in completely mastering the European languages, nor did a typical European businessman want to learn Cantonese; in any event, the latter was forbidden on pain of death. Chinese Pidgin English arose to fill the gap and allow businessmen without a common language to speak with each other, although the exact details of how it came into being are unknown. It served admirably for everyday business purposes, but it was very restricted in its domain. It was an informal language, rarely written or used for official purposes, rarely spoken between two people who knew standard English, and rarely learned by children as a native language.

If you are a typical speaker of standard English, our Chinese Pidgin dialogue probably struck you as a ridiculous, ungrammatical jumble of words. We grant that the grammar of Chinese Pidgin was simpler than that of English; there were, for example, almost no inflectional affixes, and the vocabulary was much smaller.

But Chinese Pidgin had a grammar of its own (Hall 1944). Notice that both parties in the dialog were speaking the same language; the language was not the result of a Chinese person trying but failing to speak English. If instead of “You pay he how muchee?” Hunter had said “How much will you give him?”, that would have been ungrammatical by Chinese Pidgin standards. Chinese Pidgin was a well-defined language whose vocabulary and grammar were kept very simple so that people could learn to transact business in it as efficiently as possible.

You might have expected that a contact language for bringing together Chinese and English speakers would have been a mixture of the two languages. On its surface, at least, it is overwhelmingly English. That is, in this contact situation, English served as the *LEXIFIER*, the language that furnishes most of the vocabulary. In fact, the few non-English words in our dialog, other than the proper names, happen to be not from Chinese, but from Hindi: *chop*, *mantale* < [məntri], *lac*. Below the surface, there is some Chinese influence. Many of the words have been adjusted to conform to Cantonese pronunciation patterns. The English sound [ɹ] at the beginning of a syllable usually shows up as [l], consonant clusters are often simplified (*fitty*, *sikky*), and consonants that do not appear in Cantonese syllable codas are often moved into an onset by appending [i]. A clear Chinese influence in the syntax is found in the phrase “one piece chop” (4d): *piece* is a classifier, which is required in Chinese between a number and a noun (§4.10.3.3). From the Sketch of Mandarin Chinese (§4.10) you will recognize many additional features found in most Chinese languages. For example, pronouns are invariant (e.g. the same pronoun used for both ‘I’ and ‘me’), questions use the same word order as declarative sentences, and there are no verbal inflections for number, person, or tense.

A bold summary of these observations would be that Chinese Pidgin was English vocabulary with Cantonese Chinese pronunciation, grammar, and syntax. That would suggest a hypothesis of how the language came about: Cantonese-speaking merchants were attempting to learn English. They found it relatively easy to learn English words, which are concrete and can be studied from simple lists. The more abstract principles of English grammar would be harder to learn, so they fell back on their own grammar. English, in this scenario, was the *SUPERSTRATE* language, and Cantonese was the *SUBSTRATE*. This terminology expresses the idea that English was on top of (Latin *super*) Cantonese, both in the sense of being the target that the Chinese were shooting for and in the sense that the resulting language superficially sounds more like English than Cantonese. Vocabulary tends to be more noticeable than grammar, and therefore people tend to judge Pidgin as being a variety of English, with some Chinese grammar lurking below (*sub*) the surface.

One weakness of this theory is that it ignores the English parties. If it is so hard for businessmen to learn another grammar, why did the English speakers of Chinese Pidgin manage to effectively learn Chinese grammar? Another problem is that it is not clear that all of Chinese Pidgin syntax is distinctively Chinese. It has

features not found in Chinese, such as the *-um* [əʊm] suffix (4b). More important, most of the grammatical simplifications that strike English speakers as typically Chinese are found in hundreds, perhaps thousands of other languages.

Facts like these have led to other theories about the ultimate origin of Chinese Pidgin English. One of the first ideas was that its source was an earlier language whose grammar was very much like Chinese Pidgin, but whose vocabulary was based on Portuguese. When English businessmen arrived in China, somebody (the English? the Chinese?) took that existing contact language and RELEXIFIED it to use English words instead of Portuguese words. That theory is historically appealing, because the Portuguese did precede other Europeans in establishing significant commercial activities in China. But that theory doesn't really account for many facts about the structure of Chinese Pidgin, other than the presence of a few Portuguese words. Portuguese verbs are highly inflected and therefore unlikely to be the inspiration behind the lack of inflections in Chinese Pidgin.

We lean toward the theory that the grammar of Chinese Pidgin English was not taken from any one language. Some principles, such as the basic SVO and adjective-noun word orders, were common to basic phrases in both languages and so were obvious candidates for Chinese Pidgin English. Other features deemed necessary for a language may have been pulled in from one language or the other, or even from other languages known to some of the speakers. In the early years, there was no doubt much variation as people tried introducing features. Certainly a guiding principle was always the desire to keep the language simple. People usually have a good notion about what would make a language simple, even if it would be ungrammatical in their own native language. They often talk more simply to small children or to people not fluent in the language they are using. And who of us has not at one time or another thought English would be much improved if it got rid of its irregular verbs? Such notions must have guided the development of Chinese Pidgin.

After the Opium Wars, Britain forced China to accept increased foreign trade and emigration of indentured servants. These events led to a great spread of Chinese Pidgin English both within China and abroad. The heyday of Chinese Pidgin was short-lived, however, as people who might earlier have learned Chinese Pidgin had increased opportunities and motivation to learn and use Standard English, which has higher social standing. Today Chinese Pidgin is virtually extinct, except that it probably contributed to a similar language used by a few thousand people in Nauru, an island in Micronesia. Nowadays, implying that Chinese people use Chinese Pidgin English is a highly charged racial insult.

Linguists are now aware that similar contact languages have arisen throughout the world. They are referred to as *pidgins*, a respelling of *Pigeon*. The contact that leads to the need for a language is usually afforded by trade or conquest. The Chinese Pidgin English we described above is a clear example of a pidgin that originated to facilitate trade contacts. Another example is Chinook Jargon (CHN),

or Wawa, which used to be spoken in the U.S. Pacific Northwest and adjacent regions of Canada, extending as far as Alaska and Yukon. The lexicon was based mostly on the Native American languages Chinook and Nootka. When white traders reached the area, they found this contact language useful and added many words from English and especially French, because many of these early white traders were voyageurs from French Canada. This pidgin fell out of use in the early 20th century, but quite a few enthusiasts in the Pacific Northwest are interested in revitalizing it, and loanwords from Wawa are found in many American languages in the region.

If we compare Chinook Jargon with Chinese Pidgin, we see important commonalities. Both languages have a small vocabulary; one 19th-century dictionary of Jargon had about 500 words. They have a simple grammar, with no inflection. Because the American languages that the creators of Chinook Jargon spoke had a complex, polysynthetic morphology, the fact that Chinook Jargon and Chinese Pidgin both had little or no inflection begins to indicate that pidgins in general have a basic tendency to be simple. Thus the simple morphology of Chinese Pidgin English is not necessarily due to the influence of Chinese grammar. As for the phonology of Chinook Jargon, it contains many typologically unusual sounds. It has ejective consonants, lateral fricatives and affricates, and uvular stops and fricatives. For example, the word for 'doubt' or 'uncertainty' is [t̪ʰunas]. Such sounds strike Europeans as difficult, but they are very common in the Pacific Northwest and were therefore easy for the earliest speakers of Chinook Jargon. Apparently people who develop pidgins do not always perceive typologically rare features as complexities requiring simplification.

Pidgins can also result from major upheavals. During wars and military occupations, large numbers of soldiers come into contact with other soldiers and with civilians who do not speak their language. Since numbers of trained interpreters are usually very limited, ordinary soldiers have to create makeshift means of communication, which may lead to the development of pidgins, some very short-lived. For example, during the occupation of Japan by American troops after World War II, there developed a pidgin called *Bamboo English*, used mostly by American military personnel to communicate with Japanese bar girls, bartenders, and other frequent contacts.

Pidgins, with their simplified linguistic structure, tend to be restricted in who uses them and in where they are used. But momentum toward wider use and more fixed grammar can earn them an established role in society. One such case is Nigerian Pidgin English (PCM), which is said to have arisen out of contact with the British in seventeenth-century trade, including the slave trade. Like other pidgins, Nigerian Pidgin English probably began with a small vocabulary, limiting its range of uses and its ability to express ideas with precision. Its grammatical structure would have avoided complex phrases and clauses. In today's Nigeria, which has hundreds of different ethnic languages, Nigerian Pidgin English serves as a lingua franca for the country, a common language that is widely understood across

many language communities. Its long history of use in a wide variety of contexts for many different purposes, not just trade, has led to a substantial increase in vocabulary and a more complex and stable grammar. Thus, from a grammatical perspective, it is an example of an EXPANDED PIDGIN.

Table 8.2 lists some pidgins of current and historical importance. Lingua Franca was used for centuries as a contact language in the Mediterranean, becoming the eponym for any language used for communication among several peoples that speak different languages. Today it is essentially extinct, as are Russenorsk, Chinook Jargon, and Mobilian Jargon. Pidgins rarely survive very long because they die out as soon as the social situations that necessitated them change. Occasionally, though, pidgins are so successful that they come to be used as a first language by many speakers and are passed down to children who grow up as native speakers. Current linguistic convention is to use the word *pidgin* only for languages that generally serve as a second language used for contact between speakers of different languages. As more people learn them as a first language, many authorities prefer to drop them from lists of pidgins. Of the languages in Table 8.2, this is happening especially to the English-based ones of West Africa and Papua New Guinea.

8.4 Creoles

The European expansion initiated by the Portuguese in the 1400s led to a massive exportation of European languages to other continents. In the simple case, the standard versions of languages like Portuguese, Spanish, French, and English became the standard languages in the new colonies. To this day, we regard standard

TABLE 8.2

Some pidgins

Name	Main lexifier	Location
Russenorsk	Norwegian, Russian	Norway
Gastarbeiterdeutsch	German	Germany
Chinese Pidgin English (CPI)	English	China
Cameroonian Pidgin English (WES)	English	Cameroon
Liberian Pidgin English (LIR)	English	Liberia
Nigerian Pidgin English (PCM)	English	Nigeria
Fanagalo (FNG)	Zulu	South Africa
Lingua Franca (PML)	Italian	Tunisia
Hiri Motu (HMO)	Motu (MEU)	Papua New Guinea
Tok Pisin (TPI)	English	Papua New Guinea
Chinook Jargon (CHN)	Chinook	Canada
Mobilian Jargon (MOD)	Choctaw	United States

Québec French as perhaps a distinctive accent or dialect of French, not a separate language. But there are cases in which the European language turns up in a radically different form. Consider this example of Haitian Kreyòl, which is spoken by virtually everyone in Haiti (6, our translation):

(6) Haitian Kreyòl

Gen yon ekip MIT ki reponn prezan nan efò k ap fèt pou repara dega ke
 goudougoudou 12 janvyè 2010 la te pwovoke nan inivèsite ann Ayiti. Kraze brize sa
 a te rive kreye yon opòtinite pou Ayiti: yon chans pou nou bati yon nouvo sistèm
 ansèyman ak nouvo zouti pou yon nouvo jenerasyon etidyan ak pwofesè.

gè jō ekip emajti ki upeōn pwezā nā efō k ap
 have(EXIST) INDEF.ART team MIT REL.SBJ respond present(ADJ) in effort REL IPFV

fèt pu upeape dega ke gudugudu duz
 made(PST.PTCP) PREP.PURP repair damage REL.OBJ Goudougoudou 12

zāvjē de mil dis la te pwovoke nā inivesite ān ajiti kuape
 January 2010 DEF.ART PST provoke in university in Haiti crash

buape sa a te upe kuape jō opōtinite
 break DEM DEF.ART PST happen create INDEF.ART opportunity

pu ajiti jō fās pu nu bati jō
 PREP.BEN Haiti INDEF.ART chance PREP.PURP 1PL.PRO build INDEF.ART

nuvo sistēm āsejmā ak nuvo zuti pu jō nuvo zeneuapejō
 new system education with new tool PREP.BEN INDEF.ART new generation

etidjā ak pwofese
 student with professor

‘There is an MIT team that is participating in efforts that are being made to repair damage that the Earthquake of 12 January 2010 caused in universities in Haiti. That violent destruction happened to create an opportunity for Haiti: a chance for us to build a new system of education with new tools for a new generation of students and professors.’ (MIT-Haiti 2015, *Ki sa sit sa a ye?*)

Haiti was part of the French empire until 1804, which might lead us to expect that the common language of the country would be French. The orthography of our sample text follows the official 1980 spelling reform and does not look at all French, but the words they represent sound fairly close to modern French, especially if we concentrate on the content words. Many are close or identical in pronunciation to standard French words, for example [ekip] *équipe*, [fèt] *faite*, [dega] *dégât*, [duz] *douze*, and so forth. Many others differ only in the treatment of French ⟨r⟩, which is usually [ʁ] in French but is either a velar approximant

in Kreyòl or lost entirely, for example [kɔ̃ʒe] *créer* (with a regular insertion of [j] between two vowels) and [ɔ̃ʒive] *arriver* (with a sporadic loss of the initial unstressed syllable). Another pervasive set of changes is the unrounding of front vowels, as in [etidjã] *étudiant* (French [etydjã]) and [pɔ̃fese] (*professeur* [pɔ̃fɛ sœʁ]), but many Kreyòl speakers still pronounce the original rounded vowels and decry the fact that the official orthography doesn't spell them as such (Schieffelin & Charlier Doucet 1992). Along with a spelling reform, all we have seen so far are some pronunciation differences not much greater than those that differentiate European and Canadian French.

The grammar of Haitian Kreyòl, however, differs markedly from that of standard French. Phrases like [sistem ɔ̃ʒjmã] 'system of education' show that a noun can modify another simply by being placed after it, a construction more reminiscent of Arabic (§5.6.4.1) than French, which typically uses the preposition *de* 'of'. Adjectives do not agree with their head nouns: note the invariant [nuvo] 'new' before not only [sistem] (in French masculine: *nouveau système*), but also before [ʒeneɔ̃ʒsjõ] (in French feminine: *nouvelle génération*). The word order is SVO, like French, but the definite article and the demonstrative determiner both follow the head noun instead of preceding it. Moreover, unlike in French, the demonstrative and the article can be used together: [kɔ̃ʒaze buʒize sa a] 'that destruction'. It may not be obvious from this brief passage, but verbs appear in a single invariant form, which usually resembles the French infinitive. Instead of using inflection, Kreyòl uses auxiliary verbs to indicate tense and aspect. These verbs are mostly derived from French grammatical words, but their phonology and syntactic use are very different from how they are used in French. The past is indicated by using [te] before the verb, as in [te ɔ̃ʒive] 'happened'. This verb comes from French *était* 'was', but **était arriver* would not mean 'happened' in French. The verb [ap] is placed before a verb to indicate an imperfective aspect, typically indicating an action in progress. It comes from the French preposition *après* 'after', but it would be difficult to extract an imperfective connotation from any usage of the word *après* in French.

All these differences between Kreyòl and French invite explanation. A minority perspective is that the amount of change is not shocking for two languages that have gone their own ways for more than two centuries. However, such an explanation needs to take into account the fact that standard French is an official language of Haiti, alongside Kreyòl, and is used in many social settings, including education. Given such a close contact, how could Kreyòl have wandered so far away from French? One answer lies in the social stratification of Haitian society. Over the past two centuries, there has been a great division between the upper and lower classes. The great majority of Haitians were poor peasants who had scant formal education and limited dealings with the literate, wealthy *élite*. The economic divide correlates with a racial one, the lower class being almost entirely of

African descent and the upper class being of mixed race, with some French ancestry. As one might expect, the latter cultivated literary French while the poor felt much less pressure to adhere conservatively to that standard.

Most linguists believe that Haitian Kreyòl is the result of not only gradual change over the centuries, but of an abrupt creation that violated the usual cladistic model of gradual change and transmission from parent to child. Today's social stratification is a legacy of a much greater stratification in the past: slavery. In the 1700s, Haiti was under the control of French colonists who acquired great wealth by exporting products, like sugar, that were grown by enslaved laborers who were constantly being brought from Africa. By the end of the century, the population of Haiti was at least 90% African, the rest being mostly white French colonists or their mixed-race descendants. As we saw in Chapter 5, even small areas of Africa are home to many different languages, and so it is fairly certain that the African population of Haiti had no common African language. This lack of a prior common language, combined with the necessity of engaging in some basic communication with the French colonists, inspired them to learn French. But their access to French must have been very limited. There was no universal education to teach them the standard language, and informal contact must have been meager as well: their parents didn't speak French, and extremely few Africans socialized extensively with fluent speakers of French, who were mostly white colonists and their mixed-race children. They managed to learn the basic lexical vocabulary – the content words – fairly accurately, but the grammar ended up quite different from that of standard French.

The exact process by which French turned into Kreyòl is a matter of great interest to many linguists, in part because analogous transitions happened in many other places (Table 8.3). These analogous languages are called *creoles*. The closest analogy to Haitian Kreyòl occurs when a European power set up plantations on another continent, bringing together heavily subordinated workers who spoke a variety of languages. More loosely, a creole emerges whenever a group of people construct or imperfectly learn a language and accept it as their first language, passing it on from parent to child across generations. On this definition, a creole differs from a pidgin only in that the latter is not anybody's native language.

Many theories about the origin of the grammar component of creoles are similar to theories we discussed with respect to pidgins such as Chinese Pidgin English. Even though most creoles we know about developed recently, there is, unfortunately, very little firm historical data that allows us to decide which of these theories is correct.

The *MONOGENESIS* theory, as the term is applied to creoles, holds that there was an earlier language, perhaps *Lingua Franca*, of which all creoles are basically relexifications, word-for-word translations into a different target language. This is meant to account for similarities among creoles, but other explanations are more in favor nowadays.

TABLE 8.3

Some creoles

Name	Size	Lexifier	Location
Indo-Portuguese (IDB)	4	Portuguese	Sri Lanka
Sri Lankan Creole Malay (SCI)	5	Malay	Sri Lanka
Naga Pidgin (NAG)	5	Assamese (ASM)	India
Macanese (MZS)	2	Portuguese	China
Morisyen (MFE)	6	French	Mauritius
Réunion Creole French (RCF)	6	French	Réunion
Seselwa (CRS)	5	French	Seychelles
Crioulo (POV)	6	Portuguese	Guinea-Bissau
Krio (KRI)	7	English	Sierra Leone
Kituba (KTU)	7	Kongo	Congo (DRC)
Sango (SAG)	7	Ngbandi	Central African Republic
Juba Arabic (PGA)	5	Arabic	South Sudan
Kriol (ROP)	5	English	Australia
Torres Strait Creole (TCS)	4	English	Australia
Bislama (BIS)	5	English	Vanuatu
Chavacano (CBK)	7	Spanish	Philippines
Pijin (PIS)	6	English	Solomon Islands
Hawaii Creole (HWC)	6	English	United States
Louisiana Creole (LOU)	5	French	United States
Gullah (GUL)	4	English	United States
Bahamas Creole English (BAH)	6	English	Bahamas
Jamaican Creole (JAM)	7	English	Jamaica
Haitian Kreyòl (HAT)	7	French	Haiti
Guadeloupean Kreyol (GCF)	6	French	Guadeloupe
Saint Lucian Creole French (ACF)	6	French	Saint Lucia
Trinidadian Creole English (TRF)	7	English	Trinidad
Papiamentu (PAP)	6	Portuguese	Curaçao
Belize Kriol (BZJ)	6	English	Belize
Palenquero (PLN)	3	Spanish	Colombia
Islander Creole English (ICR)	5	English	Colombia
Guyanese Creole English (GYN)	6	English	Guyana
Aukan (DJK)	5	English	Suriname
Saramaccan (SRM)	5	English	Suriname
Sranan (SRN)	6	English	Suriname
Guianese Creole French (GCR)	5	French	French Guiana

Substrate theory holds that the grammar of creoles is taken from the original language of the subordinated workers, although grammatical morphemes were not borrowed as such. Because, as with Haitian Kreyòl, so many creoles were developed by people taken in the Atlantic slave trade, similarities between many creoles could then be explained as features of languages spoken near the Atlantic coast of Africa. This idea is difficult to prove, in part because most syntactic features are

not as special as our intuitions tell us. If Kreyòl signifies tense and aspect by putting certain grammatical morphemes before the verb – a feature that strikes some creolists as very noteworthy – finding some African language out of hundreds of candidates that signifies tense or aspect by putting some morpheme before the verb is scarcely proof that Kreyòl was influenced by an African substrate. It could have been influenced by analogous uses of auxiliary verbs in French, or it could have been invented from scratch; after all, putting a qualifying word before or after the thing being qualified is perhaps the most common of all syntactic devices.

The BIOPROGRAM theory draws on Noam Chomsky's idea that children have an innate language acquisition device, which becomes inactive in adulthood. When the first generation of African Haitian adults tried to learn French, they assembled a very simple pidgin consisting of a set of content words with essentially no grammar. When their children were exposed to that speech, their language acquisition device, which is only capable of learning grammatical language, interpreted the input as a language of the most fundamental grammatical type. They automatically did such things as imposing a fixed SVO word order and developing a system of tense and aspect markers for use before verbs. All creoles should, therefore, have essentially the same grammar. It is not clear, however, that all creoles really are more similar than can be explained by theories that make simpler assumptions. And some pidgins like Tok Pisin became quite grammatical before children began to learn them as native languages. The bioprogram theory also requires assuming that all creoles develop from a preceding, usually short-lived pidgin – a reasonable hypothesis, but one not supported by much data.

Finally, the IMPERFECT LEARNING hypothesis says that the grammar of creoles is the result of an attempt to learn another language under difficult conditions. People who try to learn another language with little input will typically pick up basic vocabulary but fail to learn features that are complex and not essential to communication, such as the use of inflectional affixes. They will be likely to generalize from their own native language the idea that word order can express the relations between words in a sentence. The more a community practices using the imperfectly learned language, the more it will find it compelling to increase communicative efficiency by settling on specific word orders, conventions for marking tense and aspect, and so forth.

The imperfect learning hypothesis would seem to imply that a creole will disappear once its speakers gain increased access to the language on which the creole is based. Creoles do disappear, but that outcome is by no means certain, because the motivation to learn the standard language may diminish over time. The population may become satisfied with its perfectly adequate means of communication and may even develop a special pride in it, perhaps in defiance of the upper classes or the colonizing metropolis. Sometimes, as in Haiti, society may stay in a long-lived state of diglossia, where Kreyòl is used for certain purposes and French for others, based largely on the level of formality called for the situation; compare the similar situation found throughout the Arabic speaking world (§5.6.1). In such a

context, sociolinguists often refer to the creole as the *BASILECT* ('base language') and the standard language as the *ACROLECT* ('high language'). More subtly, the true situation is often a *CREOLE CONTINUUM*. Any given utterance can contain elements from both the basilect and the acrolect, the ratio depending on sociolinguistic factors such as the setting of the discourse, the topic, the speakers' knowledge of the acrolect, and so forth. The term *MESOLECT* ('middle language') is sometimes applied to utterances that are not on either extreme of the creole continuum.

Most of these theories predict that a creole will be simpler than the lexifier language, at least insofar as morphological complexity is concerned. This does seem to be the case, but it would be rash to extrapolate from this that all morphologically simple languages are creoles. For example, analytic languages like Chinese or Hawaiian are not clearly more complex than Haitian Kreyòl. Structurally, creoles are typical languages. The only reason they fall into a special category is the historical fact that they developed abruptly as a result of the cladistically unusual transmission between populations.

8.5 Sign languages

8.5.1 NATURE OF SIGN LANGUAGE

SIGN LANGUAGES are developed and used primarily by communities of deaf people (see Table 8.4 for the names, locations, and ISO codes of sign languages). Up to this point, we have been dealing exclusively with *ORAL LANGUAGES*, that is, languages whose words are formed in the vocal tract and transmitted as sound waves to the listener's ears. Sign languages are different. They are transmitted visually, being produced mostly by configurations and movements of the hands, arms, and face. The term *sign* is unfortunately ambiguous: linguists have another use for the term *sign* that applies equally to oral languages. To the linguist, a morpheme or a word is a *sign* that signifies an object or concept. Linguists also have a technical problem with the term *oral* in *oral language*, because sign languages too may contain oral components, such as pursing the lips or showing the tongue. A more pernicious problem with these terms is that the tradition of saying that deaf people "make signs" while hearing people "speak words" invites the error of thinking that the signs of sign language are not real words. But we defer to common usage.

In this book, we reluctantly gave precedence to oral languages. One reason is that most people communicate entirely by oral languages, and the great majority of the languages of the world are oral. To be candid, an additional reason for this precedence is that linguists know much less about sign languages than about oral ones. There have been tantalizing reports going back to at least 1200 BC of deaf people communicating by sign language (McBurney 2012: 912), but descriptions have been very sketchy; centuries later, Plato told us no more than that deaf people made signs with their hands, their heads, and the rest of their body (*Cratylus*, 422e). Reasonably informative descriptions of sign languages only

TABLE 8.4

Some sign languages

Name	Location
British Sign Language (BFI)	United Kingdom
Lyons Sign Language (LSG)	France
Adamorobe Sign Language (ADS)	Ghana
Bamako Sign Language (BOG)	Mali
Tebul Sign Language (TSY)	Mali
Hausa Sign Language (HSL)	Nigeria
Ghanaian Sign Language (GSE)	Ghana
Bura Sign Language	Nigeria
Mbour Sign Language	Senegal
Al-Sayyid Bedouin Sign Language (SYX)	Israel
Persian Sign Language (PSC)	Iran
Indo-Pakistani Sign Language (INS)	India
Chiangmai Sign Language (CSD)	Thailand
Chinese Sign Language (CSL)	China (PRC)
Hong Kong Sign Language (HKS)	China (PRC)
Taiwanese Sign Language (TSS)	China (Táiwān)
Haiphong Sign Language (HAF)	Vietnam
Hanoi Sign Language (HAB)	Vietnam
Ho Chi Minh City Sign Language (HOS)	Vietnam
Penang Sign Language (PSG)	Malaysia
Bengkala Sign Language (BQY)	Indonesia
Maritime Sign Language (NSR)	Canada
Quebec Sign Language (FCS)	Canada
Hawaiian Sign Language (HPS)	United States
American Sign Language (ASE)	United States
Plains Indian Sign Language (PSD)	United States
Yucatec Maya Sign Language (MSD)	Mexico
Jamaican Country Sign Language (JCS)	Jamaica
Providencia Sign Language (PRZ)	Colombia
Kaapor Sign Language (UKS)	Brazil

These countries have sign languages straightforwardly named after them (e.g. SQK is Albanian Sign Language): Albania SQK, Algeria ASP, Argentina AED, Armenia AEN, Australia ASF, Austria ASQ, Bolivia BVL, Brazil BZS, Chad CDS, Chile CSG, Colombia CSN, Costa Rica CSR, Croatia CSQ, Cuba CSF, Czechia CSE, Ecuador ECS, El Salvador ESN, Finland FSE, France FSL, Germany GSG, Greece GSS, Guatemala GSM, Honduras HDS, Hungary HSH, Indonesia INL, Ireland ISG, Italy ISE, Jamaica JLS, Japan JSL, Jordan JOS, Korea KVK, Lithuania LLS, Malaysia XML, Malta MDL, Mexico MFS, Mongolia MSR, Nepal NSP, New Zealand NZS, Nicaragua NCS, Nigeria NSI, Netherlands DSE, Panama LSP, Paraguay PYS, Peru PRL, Philippines PSP, Poland PSO, Portugal PSR, Puerto Rico PSL, Russia RSL, Saudi Arabia SDL, Singapore SLS, Slovakia SVK, South Africa SFS, Spain SSP, Sri Lanka SQS, Sweden SWL, Thailand TSQ, Tunisia TSE, Turkey TSM, Uruguay UGY, Venezuela VSL.

started appearing about 200 years ago, and the first modern linguistic study did not appear until 1960 (Stokoe).

In part because academia ignored sign languages for so long, several misconceptions about such languages have become widespread even among the well

educated. One misunderstanding is that sign languages consist of natural gestures and pantomime that are used in an ad hoc way to convey messages, much as in a game of charades. There is a small grain of truth in this conceptualization, in that the signs of sign languages do tend to have iconic origins. For example, in American Sign Language (henceforth ASL), a word for 'father' consists of touching the thumb to one's forehead, and extending and waving the other fingers. Part of this sign is iconic: it is made high on the face because fathers are on average taller than mothers. But this amount of iconicity is scarcely enough to enable people unfamiliar with ASL to reliably discern its meaning. Neither is it such an obvious portrayal of fatherhood that we would expect it to be uniform throughout the world. To our knowledge, no other sign language has this word for 'father', unless it borrowed it from ASL. There are many different sign languages throughout the world, each containing many hundreds of words whose signs are conventional, effectively functioning like Peircian symbols rather than icons (§2.1).

Another misconception of sign languages is that they are versions of the ambient oral language: that ASL is Indo-European, like English, or that Finnish Sign Language is Uralic, like Finnish. Such an idea would be meaningful if sign languages were recordings of oral language, unit-by-unit translations of units of oral language into visual signs. Such systems do exist: FINGER SPELLING is a system of spelling out words of an oral language letter-by-letter, and educators sometimes make pedagogical use of systems of MANUALLY CODED LANGUAGE that represent oral languages word by word (e.g. Signing Exact English, Gustason & Zawolkow 1993). And we do not deny that true sign languages can borrow vocabulary from the ambient oral language. This is often done by creating a sign based in part on one or more fingerspelled letters from the written word. For example, 'father' in British Sign Language is made by tapping the fingers twice in such a way as to represent the letter F. But sign languages that are learned and used naturally as the native language of groups of deaf speakers are independent languages that have different vocabularies and grammars from the ambient oral languages.

ASL is the most widely used and studied sign language in the world, with an estimated 250,000 users. Stokoe's pioneering study of the language (1960) was instrumental in dispelling the notion that sign languages were merely collections of gestures without any grammatical structure. Stokoe demonstrated that ASL resembled oral language in having rule-governed phonology, morphology, syntax, and semantics. The sign language can even be described with much the same terminology and theory as oral languages. For example, ASL syntax is generally described as being basically SVO, although other orders can be derived through such common syntactic processes as topicalization, whereby the topic can be moved to the beginning of the sentence even if it is the object of the verb.

At the same time, ASL, along with many other sign languages, has properties not shared with oral languages. In particular, its phonetics are understandably different because of their visual rather than aural modality. Words tend to be short, typically around three phonemes long in the analysis of Liddell & Johnson (1989),

but many phonetic features are packed into each phoneme. Each hand can take on upwards of 150 shapes, which Liddell & Johnson described in terms of combinations of 13 phonetic features. Further phonemic contrast is expressed by the location of the hands, their orientation, and their motion. A smaller but still important amount of information is also conveyed by movements of the torso, head, and face. Such nonmanual information is often intonational, that is, applying to an entire clause or sentence. For example, a polar question is signaled by raising the eyebrows and tilting the head forward. Topicalization is indicated by raising the eyebrows but tilting the head back while signing the topicalized part of the sentence. This helps interlocutors distinguish topicalized OVS sentences from more neutral SVO sentences. An appreciable number of signs use nonmanual information for lexical meaning, too. To take a vivid example, the sign for ‘to hang (a person)’ ends with the signer’s head tilted to the side, the mouth open, and the tongue hanging out.

ASL makes very little use of affixation but much use of compounding. In this respect it seems like an analytic language, perhaps a creole. However, many words of ASL combine two morphemes simultaneously, in a way that reminds one of inflectional or even polysynthetic languages. For example, many verbs can express aspectual distinctions, such as whether an action is done habitually or repeatedly, by introducing different types of motion into the sign. Many verbs adjust their signing location to spatially indicate subjects or objects of the action named by the verb. The word for ‘tell’ consists of placing the index finger on the chin and moving it away. Ideally it ends up pointing to the person to whom something is being told. Verbs of motion and location are made with the hand configured into CLASSIFIERS: a small number of shapes that broadly indicates the type of thing that is moving, thus combining the verb and a pronoun for the object into one word. Some signs incorporate numbers. The word for ‘week’ is made by drawing an index finger along the length of the palm side of the other hand. To express a certain number of weeks, the same sign is made, but instead of using an index finger, the moving hand is configured into the sign for a numeral.

A feature of ASL for which it is difficult to find a good analogy in oral languages is the way in which it makes metaphorical use of space. An object of discourse can be associated with a particular position in space, perhaps by pointing somewhere after naming or describing it; this is called INDEXING. From then on, the signer can refer to that entity simply by indicating that position, as if the person or object were actually there. The aforementioned word for ‘tell’ can take advantage of indexing by using that position as the object of the verb. That is, if the sign for ‘tell’ ends up pointing to the location previously indexed to ‘Bob’, then ‘Bob’ is the object of that verb.

8.5.2 SOURCES OF SIGN LANGUAGES

Sign languages or their precursors can develop in a variety of social situations. Through the centuries people have wondered what would happen if children were

to grow up without exposure to any language. Several cruel experiments in language deprivation have taken place, beginning with an Egyptian pharaoh's sending two children to be raised only in the company of sheep and a single shepherd sworn to silence (Herodotus 2.2). Strangely, it never occurred to these researchers that lack of exposure to oral language happens naturally all the time, when children are born deaf or become deaf before they learn their first language. Learning the ambient oral language is difficult when one cannot hear it, and, especially in isolated settings, the child may have no access to sign languages either. Deaf children who see no existing sign language in their environment tend to develop a collection of signs for communicating with their family. These HOME-SIGN systems are revealing windows onto language ACQUISITION and GLOTTOGENESIS, the creation of a language from scratch. Home-sign systems are more complex and language-like than the gestures that accompany oral language (Goldin-Meadow 2005).

Often these new sign systems spread to an entire community, forming VILLAGE SIGN. A famous case study is Al-Sayyid Bedouin Sign Language. In the middle of the twentieth century, a gene for congenital deafness began to manifest itself in an isolated Negev Bedouin village in Israel. A sign language emerged, apparently without any input from other sign languages for the first couple of generations. Sandler et al. (2005) studied eight adults, whose age meant they were of approximately the second generation to use the sign language. The researchers found a strong tendency to generate clauses in SOV order and to place numbers after the nouns they apply to. These orders could not have been borrowed from the ambient oral language, Arabic, because Arabic uses different word orders. Instead, word order preferences must have been imposed by the signing community. This is linguistically important, because using a specific word order to distinguish the relations between nouns and other elements is an element of grammar. The researchers' results indicate that Al-Sayyid Bedouin Sign Language had developed elements of grammar within a generation or two of its creation. Because grammar is a crucial property that distinguishes human language from mere collections of signs (in Peirce's sense of the word, §2.1), these findings give us an idea of how fast languages can develop when they are not based on existing languages. A related line of research is to see to what extent independently developing sign languages share the same properties in early stages of development. To take a completely hypothetical example, if 90% of sign languages turned out to have SOV word order in the second generation, one might conclude that SOV word order is more conducive to communication efficiency or more compatible with humans' innate linguistic propensities than other word orders.

Al-Sayyid Bedouin Sign Language perfectly illustrates the typical conditions for the development of village sign: a high number of deaf people growing up in a tightly knit community, especially if that situation continues across multiple generations. Typically all the deaf people in the village learn the same sign language, often from an early age. So do many hearing people. In fact, because any single deaf person will typically have many hearing family, friends, and colleagues to

communicate with, often the number of hearing people who know the sign language appreciably exceeds the number of deaf users. In Adamorobe in Ghana, it has been reported that in the recent past, about 10% of the villagers were deaf, but the village sign language was used to at least some extent by most of the population. Other village sign languages include Bura Sign Language of Kukurpu village in Nigeria, Tebul Sign Language in the village of Uluban in Mali, Yucatec Maya Sign Language in Mexico, Jamaican Country Sign Language in Saint Elizabeth parish, and Providencia Sign Language in Colombia. Martha's Vineyard Sign Language (MRE) is perhaps the most famous historical village sign language. It developed in the 18th century in and around the town of Chilmark in Massachusetts, which used to have a very high rate of inherited deafness. A village sign language that sprang up in Bengkulu, Indonesia, has 1,200 users, of whom 50 are deaf.

DEAF-COMMUNITY SIGN languages arise when deaf people congregate regularly over an extended period of time. In an urban environment, deaf people who socialize with each other may develop a sign language to facilitate communication. Little is known about how these languages form, but it is generally assumed that the process involves some synthesis or compromise among the home-sign systems and village sign languages that individuals bring to their new community. We have read about urban community sign languages in places like Bamako in Mali, Kano in Nigeria (Hausa Sign Language), and Mbour in Senegal.

Deaf-community sign languages also form in schools, where their development may be influenced by sign languages introduced by teachers, who often come from or train abroad. It is common to attribute school languages to the teachers, and they may indeed have a strong influence, especially on the languages' vocabulary. But in many cases the main source of a school language is provided by the students, perhaps from one or more village or deaf-community languages they already knew. One of the most influential sign languages in the world, French Sign Language, originated at a school for deaf people that opened in 1760. Traditionally, the school's founder, Abbé Charles-Michel de l'Épée, has been credited with inventing that language. But de l'Épée's own pedagogical writings described a very laborious system of manually coded French that is not at all like French Sign Language. A reasonable hypothesis is that it was his students who supplied the sign language, possibly one already used by a deaf urban community in Paris. Even when the established French Sign Language was brought to other communities – schools for deaf children sprouted up across Europe and North America, often staffed by teachers trained at de l'Épée's school – it must have been mixed with local sign languages supplied by the deaf students. ASL, for example, developed when Laurent Clerc, who had been trained in de l'Épée's school, taught French Sign Language to his students in Connecticut. But contemporaneous documentation is scanty, and we may never know how much of the ASL is due to French Sign Language and how much may be due to students, especially those already familiar with other languages, such as the village sign of nearby Martha's Vineyard.

Particularly strong evidence that school languages can be introduced by students rather than teachers comes from the fact that they spring up even in rigidly ORALIST schools: those that neither teach nor even allow the use of sign language, out of fear that using sign would take time and resources away from practicing oral language. Nicaraguan Sign Language (NCS) started forming in Managua when the first of two strictly oralist schools for deaf children opened there in 1977. In the 1980s, linguists were already on the scene to study the new deaf-community sign language that the children developed. Observing different cohorts of graduates of those schools gave linguists an opportunity to study how this new language evolved. It also gave theorists an opportunity to argue about whether this case of glottogenesis proved Chomsky's theory about a language acquisition device. One expectation was that the communication system devised by the first cohort could lack essential properties of human language, but that when subsequent cohorts of children learned that existing system as if it were a language, their language acquisition device would turn it into a normal human language. It does appear that Nicaraguan Sign Language became more complex and rule governed over the years, but not everyone was convinced that Chomsky's theory was the only possible way to account for that fact.

A major research topic in sign language typology concerns how deaf-community sign languages – whether originating in cities or schools – differ from village sign languages and home sign systems. One interesting observation is that only deaf-community languages appear to exploit space in a grammatical way. Such properties have not been reported in village sign (Meir et al. 2010), possibly because the oral languages spoken by many village signers exert a conservative influence, rejecting innovations that are too dissimilar from how oral languages work.

Less attention has been paid to sign languages that are not oriented to the needs of deaf people. ALTERNATIVE SIGN languages are common in many Australian Aboriginal communities, especially those in the central desert region. An alternative sign language is a manual communication system that is used primarily by speakers of an oral language who find themselves in a situation where sounded language is ineffective or prohibited. Hunters often develop a basic vocabulary of signs for communicating with each other without alerting their targets, and such a goal may have been behind the initial development of sign languages among the hunter-gatherers of Australia. An additional factor is that many Aboriginal communities have traditions discouraging vocal speech in certain situations, especially those calling for respect or mourning. Groups such as the Warlpiri go so far as to ban widows from using vocal speech for a year or more after their husbands die, an exigency that has led to the development of a set of manual signs that is as complete as the vocabulary of the oral language. Amazing as these systems are, linguists tend not to call them distinct languages in their own right. Their vocabulary tends to map word for word, often morpheme for morpheme, onto the oral language, and their grammar mimics that

of the oral language or just omits elements such as inflections that can usually be inferred from context. That is, from a structural point of view, signed Warlpiri is considered to be the same language as oral Warlpiri, a manually coded form of it (Kendon 1988).

In other parts of the world, manually coded languages have been used by deaf educators in an attempt to visually familiarize deaf pupils with the structure of spoken language. More in the spirit of the Australian sign systems, monastic communities in Europe and elsewhere that forbid vocal communication at certain times, most notably the Benedictines and Cistercians, have developed sets of manual signs that number in the hundreds. These systems too are not normally regarded as true languages *per se*, both because of their limited vocabulary and because they do not have their own systems of grammar.

Plains Indian Sign Language (Davis 2010) was a highly developed system of thousands of manual signs that was used as an alternative language by some three dozen Indian nations who lived in the Great Plains of North America, including those who spoke Cheyenne and Arapaho. Evidently, its primary use was communication between people who spoke different languages, arguably making it the most sophisticated signed *lingua franca* known.

8.5.3 SPREAD OF SIGN LANGUAGES

Even our short discussion of different types of sign languages may have given you some notion about why a cladistic analysis of sets of sign languages is at best difficult, at worst ill-defined. New sign languages sprout up spontaneously when deafness arises in a home or village. When deaf communities form in cities or schools, a new language may be formed by combining parts of two or more existing languages. To complicate matters even more, some sign languages are constructed by speakers of oral languages. Thus sign languages combine all the cladistics-busting problems of pidgins, creoles, mixed languages, and constructed languages. It is therefore safest to not even try to speak in cladistic terms, about families and clades of sign languages, but rather to discuss how those languages spread.

Even that simpler charge proves to be rather difficult. You might think that, because most known sign languages presumably originated in modern times, the historical record would be a great help. Unfortunately, the historical record contains little information about sign languages: very few lexicons and essentially no information about grammar. To a large extent, we are left to speculate about how home sign, village sign, and urban sign languages form and, especially, how they interact. It does seem to be the case that village sign is often an expanded version of home sign; indeed, there may be no real distinction in villages where most people are related to each other. Yucatec Maya Sign Language, for example, spread from the village of Chican, and Martha's Vineyard Sign Language spread to the entire island from the town of Chilmark.

As our discussion of creole languages may have suggested, colonialism can play an important role in spreading languages. For example, British Sign Language spread throughout the United Kingdom and several colonies, giving rise to sign languages in Australia, New Zealand, and South Africa, and also to a Maritime Sign Language in Canada. French Sign Language, or its derivatives, have been reported in Algeria, Congo (DRC), Mali, Rwanda, and Togo. Italian Sign Language or its derivatives have been reported in Ethiopia, Libya, and Tunisia. Japanese Sign Language spread to Korea and Taiwan. But education systems can play an important role, too. De l'Épée's success in educating deaf students inspired many schools to be founded in other countries, using his methods. Consequently, French Sign Language spread to, or contributed significantly to, sign languages in Austria, Hungary, the Czech Republic, Slovakia, Croatia, and other former Yugoslavian nations, as well as Russia and its former client states. As we mentioned earlier, it also contributed to ASL, which spread rapidly through the United States by the agency of dozens of residential schools. It or a descendant language is also widely used in Canada and certain other American countries such as Barbados and Bolivia, in Malaysia and Indonesia, the Philippines, and several other countries. Missionary ambitions can also be a factor in the founding of schools that spread ASL. Starting in 1956, the deaf African American missionary Andrew Foster opened 31 schools in 13 different African countries, primarily in Western and Central Africa. His schools practiced TOTAL COMMUNICATION, a philosophy that incorporates sign language as well as training in oral language. A graduate of Gallaudet College (now Gallaudet University), Foster used ASL. Consequently that language or a descendant is now used in Ghana, Nigeria, Chad, and many other countries.

The wide dissemination of ASL makes it something of a lingua franca in the global deaf community. Gallaudet University, which is bilingual in ASL and English, is one of very few institutions of higher education where all classes are accessible via sign language, making it a magnet for international deaf students and making ASL an important language of post-secondary learning.

8.6 Language endangerment and renewal

Languages can disappear suddenly or gradually. Natural catastrophes or genocide campaigns have been known to kill off a language along with all its speakers. For example, in Indonesia the Tambora volcano erupted unexpectedly and violently, killing 10,000 people immediately and another 70,000 soon afterward due to starvation and disease. In the process, the Tamboran language lost all of its speakers. A language that became extinct as a result of genocide is Ubykh, a language of the Caucasus whose 50,000 speakers were expelled from Russia in the 1860s. Many speakers died while fleeing their homeland, and the rest scattered to different parts of Turkey. The last speaker died in 1992.

Most language death is not so abrupt or violent. More often, a language ceases to exist when its speakers abandon it for another, a phenomenon known as language SHIFT. When two languages come into contact, political, social, or economic factors can make one of the two dominant. Over time, the nondominant language may be abandoned, often due to active suppression by forces favoring the dominant one. On occasion, the abandonment may be abrupt. For example, in response to a peasant rebellion against wealthy landowners in 1932, the government of El Salvador massacred over 10,000 – possibly several tens of thousands – indigenous people. To avoid death at the hands of the military, the indigenous population soon abandoned markers of their cultural identity, including their languages. In another famous case, all of the indigenous languages of Tasmania disappeared just 73 years after the first contact with Europeans (Crowley & Dixon 1981: 396).

Typically, though, language shift happens gradually. How gradually depends on factors that include politics, educational policies, and community efforts to shore up a dying language. After the Normans conquered England in 1066, the French language dominated in many domains of life, including government and religion. But the policies in force allowed English to continue as the first language of most of the population, and eventually the official functions of French were eliminated entirely. The centuries of contact with French had lasting effects on English – including vast numbers of loanwords – but at no point was the English language threatened with extinction.

Nowadays the most common and most enduring force behind language shift is the ascendance of GLOBAL LANGUAGES. For many decades, the use of English has been expanding around the world, with significant but smaller expansions on the part of Chinese, French, and others. In Africa, a few widely used languages like Swahili in East Africa are gradually replacing smaller languages. In Tanzania, Swahili is the medium of instruction in all but a handful of public schools. Competing with Swahili is English, a compulsory subject in Tanzania in pre-school, primary school, and secondary school, and the most common medium of instruction at university level. These language policies reveal the relative status today of Tanzania's various languages and establish a self-fulfilling prophesy for the future. The situation in Tanzania reflects a common pattern worldwide. Out of about 7,000 living languages, just 100 cover 90% of the world's population. The indigenous languages of the Americas provide examples of how language shift operates. Most have been in decline for centuries, and many have become extinct. Of the rest, the vast majority are in danger, with just a small number of speakers growing up with them as a first language. Despite this, efforts continue to revitalize certain languages with education programs, community-wide awareness campaigns, and legal measures.

How worthwhile it is to try to revitalize endangered languages? Does a dying language matter, and how effective are measures to preserve such a language? Any human language embodies a distinct culture and thus includes achievements in art, science, medicine, and other forms of human knowledge, developed and applied

over many generations to solve life's problems. As Hinton (2001) also reminds us, language retention can be a human rights issue, since language preservation often goes hand in hand with efforts to achieve political recognition and, with it, a sense of one's own identity.

Adding to the case for endangered language revitalization, UNESCO has connected linguistic diversity with biodiversity. Over time, languages develop detailed classification systems for their habitat, which can be irretrievably lost when a language dies. Conversely, the loss or degradation of natural habitats adversely impacts human cultures, including their languages. For example, the Karuk in northern California are currently engaged in two preservation battles. The Karuk Language Restoration Committee is fighting to preserve the language, which now has no more than 10 fluent native speakers, and the Karuk are fighting to reverse government-sanctioned pollution that threatens salmon, which are essential to their livelihood.

Communities around the world are currently working to revitalize endangered languages. Efforts, often with outside public or private support, take many forms: school-based programs, group activities that encourage adults to engage with one another and with children in the endangered language, and efforts by specialists to develop reading materials for speakers.

The United States reached a turning point in government policy with the Native American Languages Act of 1990 (25 USC §§2901–2906, 2015), which established as official policy the right of Native Americans to use, practice, and develop Native American languages. In the United States and Canada, many tribes are using techniques specifically designed for spreading the use of native languages. At this writing, 16 states have such schools and programs, and the number of states and programs is on the rise.

While linguists have been writing grammatical descriptions of languages for centuries, one estimate is that only 10% of the world's languages are well documented. Many times in this book we have pointed to the lack of data to resolve one question or another. With the growing pace of language endangerment, language DOCUMENTATION has become a major focus of activity in the twenty-first century. Harnessing the latest technologies, computer scientists and linguists are developing more efficient and easy-to-use tools for recording much more language data from native speakers and for translating, transcribing, annotating, archiving, and distributing that data. Documentation of a language is always a useful activity, but it will become invaluable if the language loses native speakers. Such information will help linguists to develop a more complete understanding of how human languages work. Vastly improved language documentation will also aid efforts to revitalize moribund languages and revive those that no longer have native speakers.

8.6.1 EXAMPLE: AUSTRALIA AND NEW ZEALAND

Of the 390 languages in Ethnologue's listing for Australia, 171 are extinct, 141 are dying, and 35 are "in trouble". Another 18 languages have a standardized alphabet

and some written literature but are deemed not sustainable in the long run. In most schools, the only language of instruction is English. In New Zealand, the picture is different, perhaps because Maori is the only indigenous tongue. It is spoken or at least understood by most ethnic Maoris and is widely written. It can be used in legal proceedings and is the language of instruction in several hundred schools. Still, the status of Maori is sliding. It is disfavored by young speakers and is rated by Ethnologue as threatened. One dialect, Moriori, is already extinct.

One can easily imagine why adult attitudes toward the language might not be shared by youth, for whom getting ahead in the world means embracing English. This captures the quandary faced by many of the world's young people, who need to acquire a world language for career and social advancement without turning their backs on their language background and ethnic identity.

8.6.2 EXAMPLE: NAVAJO

Language shift has applied to every Native American language. Of these, Navajo has a better chance of avoiding extinction than most of the others. It has about 171,000 speakers, more than any other Native American language in the United States. Out of such a large population, the Navajos form a relatively close-knit group, with about half of them living in the Navajo Nation. For a long time, restrictions against the use of the Navajo language in schools contributed to the same sort of decline experienced by other Native American languages. However, in the latter half of the 20th century, measures by the United States and local governments to foster the use of minority languages in schooling, along with other revitalization efforts, has turned this situation around somewhat. Another healthy sign is that Navajo is used in some local broadcast media.

Like some other Indian nations, the Navajo Nation also tried imposing a requirement that the president of the Navajo Nation must be fluent in Navajo. In the 2014 election, one candidate who entered the race for the presidency was disqualified because his command of Navajo was not good enough. In 2015, however, a referendum removed this requirement. While many voters admired the intent to promote the Navajo language, many also were concerned that the requirement excessively restricted the pool of candidates who could run for president. Commentators pointed out that younger people, in particular, tended to lack fluency in the language (Fonseca 2015).

8.6.3 EXAMPLE: CELTIC

The Celtic group of Indo-European languages once extended across western and central Europe, but it is now confined to small locations around the western reaches of France, Wales, Ireland, and Scotland, having largely been displaced by languages from the Germanic and Romance groups. The smallest Celtic languages, Manx and Cornish, were regarded as extinct, but due to revitalization efforts Manx now has 1,689

speakers and Cornish about 2,000. Both are classified as critically endangered in the 2010 version of UNESCO's *Atlas of the world's languages in danger* (Moseley 2010).

The remaining Celtic languages are all in some degree of danger. The healthiest is Welsh, classified by UNESCO as vulnerable. Modern Irish, though protected by the government of Ireland, is used regularly by just a small fraction of the population, possibly 3% or less. As such, it is classified as definitely endangered, as is Scottish Gaelic. The remaining Celtic language, Breton, spoken in the Brittany district of France, is one step further along in endangerment and is classified as severely endangered.

8.6.4 CONCLUSION

Languages are more than vehicles of communication. The language we learn at home is most often a key part of our cultural identity, and when a language dies, important parts of the culture associated with it ordinarily die as well. When writing this chapter, we searched Google for the phrase "Our language is holy" and on the first page of results found this precise sequence of words applied – apparently independently – to three languages: Hebrew, Navajo, and Latvian.

Languages are sets of symbols. But in addition, a language as a whole can come to serve as a symbol of ethnic heritage, group cohesiveness, and personal identity. Whether our language is small or large, endangered or not, confined to a small area or dispersed around the globe, it likely embodies information and lore passed down through the ages. When languages are actively suppressed – as in the many cases in the developing world where use of the mother tongue is prohibited in schools – this easily becomes a human rights issue.

On the other hand, it would be hard to fault young people who choose to use a more prestigious tongue over their household language as a means of getting ahead in the world. Different languages – and different registers of the same language – have uses all their own. The truly tragic event is when the choice of language is imposed from outside. As Ash et al. (2001: 19) put it, "We do not exist in a condition of economic justice where people who choose to do so can speak a local language and pass it on to their children entirely without regard for any economic consequences. . . . The pressure to use the dominant language, and even to abandon one's local language, is quite generally overwhelming and virtually irresistible."

8.7 Sketch of Tok Pisin

8.7.1 GENERAL BACKGROUND

Tok Pisin is an expanded stable pidgin language that is in the process of acquiring native speakers. Ethnologue estimates that there are 122,000 first-language speakers, of whom 50,000 are monolingual. As the term is defined by most linguists, Tok Pisin, despite its name ('pidgin language'), has transitioned to creole status. It is now one of the official languages, along with English and Hiri Motu, of Papua New Guinea,

a country that contains a multitude of tribal languages spoken by a few scattered villages each. It is estimated that as many as four million people use Tok Pisin as a second language. A number of newspapers and magazines and an increasing number of books are published in it, and a good deal of official communications and even debates in parliament are conducted in the language. Without Tok Pisin, in spite of its humble origins as a contact language, it probably would be very difficult to govern the country. It would probably cause much strife and unrest if one of the native languages of the country were to be chosen as the official language, and to make all or a dozen languages official would encourage chaos and regional separatism.

Although the bulk of Tok Pisin vocabulary comes from English, a few words are German in origin and many others were taken from the various languages of Papua New Guinea, especially Tolai (KSD), an Austronesian language. The words from German were incorporated during the time of German influence and colonization in Papua New Guinea. However, most German words have now been replaced by borrowings from English:

- (7) a. [tinte] < German *Tinte* 'ink' (now [in] < English *ink*)
 b. [karaide] < German *Kreide* 'chalk' (now [sak] < English *chalk*)
 c. [kartopel] < German *Kartoffel* 'potato' (now [poteto] < English *potato*)

From Tolai and related languages come such words as:

- (8) a. [balus] 'pigeon'
 b. [liklik] 'small'
 c. [tamboran] 'ghost'

There is also a difference in vocabulary, and very likely some difference in grammar, between rural Tok Pisin and the urban Tok Pisin spoken in Port Moresby and a few other large centers of population, where English influence is rather strong. Thus, for example, in urban Tok Pisin the words for 'jet plane', 'orange', and 'lemon' are direct loans from English, whereas in rural Tok Pisin their equivalents are not:

- | | | | |
|-----|-------------|---------|---------------------------------|
| (9) | Gloss | Urban | Rural |
| | 'jet plane' | setplen | smokbalus (lit. 'smoke pigeon') |
| | 'orange' | orins | switmuli (lit. 'sweet lime') |
| | 'lemon' | lemen | solmuli (lit. 'sour lime') |

Tok Pisin, like most contact languages, has an analytic morphology. Its basic word order is SVO. Non-English elements involve the presence of an inclusive-exclusive contrast in the pronouns and morphological marking of transitivity in verbs, both of which features are also found in the indigenous languages of the area where Tok Pisin was formed.

8.7.2 PHONETICS, PHONOLOGY, AND ORTHOGRAPHY

The phonological system of Tok Pisin is influenced greatly by both English and the Austronesian languages of Papua New Guinea.

8.7.2.1 Consonants

Table 8.5 shows the consonant phonemes found in virtually all varieties of Tok Pisin. The language also has the glides [j] ⟨y⟩ and [w].

English-influenced, urban speech may have additional consonant phonemes, especially at the beginning of words. The postalveolar affricate [dʒ] ⟨j⟩ is used by many speakers at the beginning of words; otherwise it is replaced by [s]: [bris] ‘bridge’, [dʒas] or [sas] ‘judge’. The consonant [f] is regularly borrowed as [p], but many speakers produce a bilabial fricative [ɸ] word-initially.

The voiceless stops of Tok Pisin are unaspirated in all positions. In some dialects, the voiced stops are strongly prenasalized, due to the influence of the local Austronesian and Papuan languages. The rhotic sound can be a trill but is most commonly a tap, [ɾ].

Obstruents are devoiced in word-final position. This is reflected in the orthography, even though this devoicing is completely automatic. According to Mihalic (1971: 8), the only exception to this rule is the word *God*.

The velar [ŋ] is written ⟨n⟩ when it stands before a velar stop: *anka* [aŋka] ‘anchor’. Thus, the spelling *ng* is ambiguous, just as in English: [ŋ] or [ŋg].

8.7.2.2 Vowels

The vowel system of Tok Pisin is relatively simple. It consists of five monophthong vowels: [i], [u], [e], [o], and [a]. It also has the same three diphthongs as English:

- (10) a. [ai] as in [taim] ‘time’
 b. [au] as in [maus] ‘mouth’
 c. [oi] as in [boi] ‘boy’

The vowel phonemes have tense allophones in open syllables and lax allophones in closed syllables. The tense allophones are not pronounced with an

TABLE 8.5
Tok Pisin consonants

Stop	p	t	k
· +voice	b	d	g
Fricative	v	s	h
Nasal	m	n	ŋ ⟨ng⟩
Lateral		l	
Rhotic		r	

offglide as in English. Thus the tense allophone of the phoneme /o/ is pronounced [o], not [ou].

8.7.3 MORPHOLOGY

8.7.3.1 Pronouns

The personal pronouns are shown in Table 8.6. Note that there are no gender differences in the third person pronoun: [em] can stand for ‘he’, ‘she’, or ‘it’.

Third person subject pronouns are usually followed by the third person agreement particle [i], which comes between the pronoun and the verb (11a). No such particle exists for other persons (11b). However, [i] also appears in some constructions after other persons. See the section on the verb morphology below (§8.7.3.3). Also, some dialects of Tok Pisin do use this particle after other personal pronouns.

- (11) a. em i luk-im mi
 3 3.SBJ look-TR 1
 ‘He looks at me.’

- b. mi luk-im em
 1 look-TR 3
 ‘I look at him.’

Reflexive pronouns are the personal pronouns followed by the particle [jet]:

- (12) mi pait-im mi jet
 1 hit-TR 1 REFL
 ‘I hit myself.’

Emphatic forms of the personal pronouns are built from the personal pronouns by means of various particles:

- (13) a. [mi jet] ‘I myself’
 b. [mi tasol] ‘only I, only me’
 c. [mi wan-pela] ‘I alone, me alone’

TABLE 8.6
Tok Pisin personal pronouns

Person	SG	PL
1	mi	EXCL: mi-pela INCL: jumi
2	ju	ju-pela
3	em	(em) ol

Numbers are usually attached to the personal pronouns as suffixes:

- (14) a. [mi-tu-pela] ([mitla] in fast speech) ‘two of us’
 b. [mi-tri-pela] ‘three of us’

8.7.3.2 Adjectives and adverbs

Tok Pisin adjectives usually end in the suffix [-pela] and most precede the nouns they modify:

- (15) a. [bik-pela maunten] big-ADJ mountain
 b. [jan-pela meri] young-ADJ woman

When used predicatively, most adjectives retain the suffix [-pela] (16a–b), but some do not (16c):

- (16) a. [maunten i bik-pela] ‘the mountain is big’
 b. [dis-pela haus i gut-pela] ‘this house is good’
 c. [dis-pela haus i klin] ‘this house is clean’

A few adjectives change meaning depending on whether they have the suffix or not:

- (17) a. [dis-pela meri i stret] ‘this woman is honest’
 b. [dis-pela meri i stret-pela] ‘this woman has a straight posture’

A small number of adjectives, mostly of non-English origin, never take the suffix [-pela]:

- (18) a. [liklik] ‘small’
 b. [lapun] ‘old’

Some adjectives follow the nouns they modify:

- (19) a. [man natiŋ] ‘just an ordinary person’
 b. [ples tambu] ‘forbidden place’
 c. [samtiŋ nogut] ‘something bad’

Comparison of adjectives and adverbs is signaled not by affixation but by the word [moa] (< *more*) and various circumlocutions:

- (20) a. dis-pela moran i loŋ-pela moa loŋ dis-pela
 this-ADJ python 3.SBJ long-ADJ more PREP this-ADJ
 ‘This python is longer than this one.’
 [long] is a preposition of a multitude of meanings; here, ‘than.’

b. tasol dis-pela moran i loŋ-pela loŋ olgeta
 but this-ADJ python 3.SBJ long-ADJ PREP all

‘But this python is longest of all.’

c. em i dig-im boret hariap loŋ mi
 3 3.SBJ dig-TR trench fast PREP 1

‘He dug the trench faster than me.’

d. em i dig-im boret hariap olsem mi
 3 3.SBJ dig-TR trench fast like 1

‘He dug the trench as fast as me.’

e. em i dig-im boret hariap winim olgeta
 3 3.SBJ dig-TR trench fast SUPL all

‘He dug the trench the fastest.’

[winim] is a marker of the superlative.

8.7.3.3 Verbs

Verbs in Tok Pisin that are not marked for tense can refer to present, past, or future, depending on the context. Thus, for example, [mi go loŋ taun] may mean ‘I am going to town’, ‘I went to town’, or ‘I shall go to town’. Past tense may be marked by preposing the particle [bin]:

(21) ol kakaruk i bin ranawe i go loŋ bus
 PL chicken 3.SBJ PST run_away 3.SBJ go PREP bush

‘The chickens ran away into the bush.’

The future tense may be marked by a preposed particle [bai], which is usually shortened to [ba] or even [b] in fast speech:

(22) bai mi go loŋ taun
 FUT 1 go PREP TOWN

‘I will go to town.’

[mi bai go loŋ taun] is also possible.

The future particle usually precedes short subject noun phrases, but it usually follows long ones. In the latter case, the particle [i], which normally shows third person agreement with the subject, may follow it:

(23) mi bai i go loŋ taun
 1 FUT SBJ go PREP TOWN

‘I shall go to town.’

Transitive verbs usually, but not always, have the suffix [-im] attached to them. This suffix may be used to derive transitive verbs from intransitive ones or other parts of speech:

- (24) a. [bos] 'boss' (noun) vs. [bos-im] 'to boss, oversee'
 b. [wok] 'work' vs. [wok-im] 'to make, build'
 c. [orait] 'okay' vs. [orait-im] 'to fix up'
 d. [bihain] 'later' vs. [bihain-im] 'to follow'
 e. [op] 'to be open' vs. [op-im] 'to open'
 f. [kros] 'to be angry' vs. [kros-im] 'to scold'
 g. [siɲaut loŋ] 'to call out to' vs. [siɲaut-im] 'to call'

There are, however, a fairly large number of transitive verbs that do not have the transitive suffix [-im] attached:

- (25) a. [kaikai] 'eat' (but some people use [kaikai-m] instead)
 b. [gat] 'have'
 c. [save] 'know'
 d. mi pekpek wara
 1 defecate liquid
 'I have diarrhea.'

Various particles are used to mark aspectual and modal differences. The perfective aspect is marked by the particle [pinis], which is added after the verb or even after the object:

- (26) a. ol-pela pat biloŋ ka i bagarap pinis
 old-ADJ part of car 3.SBJ wear_out PFV
 'The second hand parts of the car have worn out.'
- b. mi rit-im wantok niuspepa pinis
 1 read-TR Wantok newspaper PFV
 'I have finished reading the Wantok.'

The progressive aspect is marked by postposing [i stap] after the verb. Note that [stap] is a general location verb in Tok Pisin meaning something like 'stay', 'be located'.

- (27) mi raun i stap biloŋ pain-im wok
 1 go_around SBJ PROG GEN find-TR work
 'I am walking around looking for work.'

Habitual action is expressed by [save] ('know' used as an auxiliary) before the verb:

- (28) em i save kaikai ol man
 3 3.SBJ HAB eat PL person
 'It used to eat people.'

'About to' is rendered by either [laik] or [klostu] before the verb. Finally, physical ability is expressed by [inap] before the verb:

- (29) mi bai inap wok
 1 FUT be_able work
 'I shall be able to work.'

8.7.4 SYNTAX

8.7.4.1 Word order

The basic word order is SVO, as in English. As already noted, most adjectives precede the nouns they modify, but there are some that follow what they modify.

8.7.4.2 Grammatical relations

Grammatical relations in Tok Pisin are marked entirely by word order and the prepositions, because even pronouns do not have different forms when they are used as subjects and objects of verbs.

8.7.4.3 Relative clauses

There are three different types of relative clause constructions in Tok Pisin. Relative clauses may occur without any overt relativizer:

- (30) dis-pela man i kam asde em i papa bilonj mi
 this-ADJ man 3.SBJ come yesterday 3 3.SBJ father GEN 1
 'This man who came yesterday is my father.'

Relative clauses may be bracketed by the demonstrative [ja]:

- (31) dis-pela man ja em i stap lonj bus ja em i redi na
 this-ADJ man DEM 3 3.SBJ live PREP bush DEM 3 3.SBJ ready and
 em i kisim bonara
 3 3.SBJ get bow_and_arrows
 'This man who lived in the bush was ready to get his bow and arrows.'

The third type of relative construction employs relative pronouns. It is found in the varieties of Tok Pisin that have been much influenced by English:

- (32) a. pablik seven em i man husat i gat stroŋ loŋ wok biloŋ en
 public servant 3 3.SBJ man who 3.SBJ have hard PREP work GEN 3
 'A public servant is someone who has a hard job.'
- b. ol i go loŋ wan-pela ples we i gat bik-pela tais loŋ en
 they 3.SBJ go PREP a-ADJ place where 3.SBJ have big-ADJ swamp PREP 3
 'They went to a place where there is a big swamp.'

The relative clauses of this type are more common where the relative pronouns have the English equivalents 'whoever' or 'whatever' rather than 'who' or 'which':

- (33) husat ol man i laik kam i ken i kam
 who PL man 3.SBJ want come 3.SBJ may 3.SBJ come
 'Whoever wants to come can come.'

8.7.4.4 Direct and indirect speech

Tok Pisin uses the marker [olsem] 'like' to introduce the indirect quotation. Compare direct speech in (34a) with its indirect counterpart in (34b).

- (34) a. em i tok mi bagarap pinis
 3 3.SBJ say 1 tired PFV
 'He said, "I am exhausted."'
- b. em i tok olsem em i bagarap pinis
 3 3.SBJ say that 3 3.SBJ tired PFV
 'He said that he was exhausted.'

8.7.4.5 Interrogative sentences

Polar questions differ from their non-interrogative equivalents only by intonation. Content questions have interrogative words such as [husat] 'who' and [wanem] 'what', 'which'.

8.7.5 SAMPLE TEXT

We are indebted to Gillian Sankoff for providing the following story. She recorded it in July 1971 in Papua New Guinea, annotated it, and explained it to author Lyovin in August 1992. It is a folktale told by an informant who is identified as Lina Z. She is a native speaker of Tok Pisin, which explains why the text contains

some complicated syntactic structures that are usually absent from the speech of second-language speakers.

As in the case of the story cited in the sketch of Dyrbal, this is not a polished literary piece but an oral presentation transcribed by a field linguist. Sankoff removed some hesitations and normalized some of the spelling, and we made a few changes on the advice of Craig Volker, then a graduate student in linguistics at the University of Hawai'i who has done field work in Papua New Guinea. Otherwise, the story is cited as it was recorded. Therefore, some of the usages do not always conform with Standard Tok Pisin. The idiomatic English translation was provided by Sankoff.

- (35) Long taim bifo, wanpela ailan, draipela pik i save stap ya, na em i save kaikai ol man.

loŋ taim bifo wan-pela ailan drai-pela pik i save stap ja
 PREP time past one-ADJ island huge-ADJ pig 3.SBJ HAB live DEM
 na em i save kaikai ol man
 and 3 3.SBJ HAB eat PL person

'Once upon a time, there was a certain island where a huge pig used to live, and it used to eat people.'

- (36) Em i save kaikai ol man nau; wanpela taim, wanpela taim, nau ol man go tokim bikpela man bilong ol, bos bilong ol, ol i go tokim em nau, em i tok: "Orait yumi mas painim nupela ailan."

em i save kaikai ol man nau wan-pela taim wan-pela taim nau ol man
 3 3.SBJ HAB eat PL person now one-ADJ time one-ADJ time now PL person
 go tok-im bik-pela man bilonŋ ol bos bilonŋ ol ol i go tok-im
 go talk-TR big-ADJ person GEN 3PL leader GEN 3PL 3PL 3.SBJ go talk-TR
 em nau em i tok orait jumi mas pain-im nu-pela ailan
 3 now 3 3.SBJ say all_right 1PL.INCL must find-TR new-ADJ island

'Since it used to eat people, one time, the people went and talked to their big man, their leader, they went and talked to him. He said: "All right, we must find a new island."'

[nau] 'now' may be translated as 'then' in narratives about past events. According to Sankoff, [nau] here is a marker of punctual aspect.

[tokim] takes a direct object in Tok Pisin, unlike its English counterpart *talk*, which requires the preposition *to*.

- (37) Nau, ol i stretim ol samting bilong ol na i go painim nupela ailan.

nau ol i stret-im ol samting bilonŋ ol na i go pain-im nu-pela ailan
 now 3PL 3.SBJ gather-TR PL thing GEN 3PL and 3.SBJ go find-TR new-ADJ island

'Then, they got their belongings together and went to seek a new island.'

- (38) Na wanpela meri, pik, pik ya bin kaikai man bilong en bifo na, em i gat bel.

na wan-pela meri pik pik ja bin kaikai man bilonŋ en bifo
 and one-ADJ woman pig pig DEM PST eat husband GEN 3.OBL PST

na em i gat bel
and 3 3.SBJ have belly

'And one woman whose husband had been eaten by the pig was pregnant.'

[ja] is a demonstrative marker that refers to items that have been mentioned before. Thus, it can be translated as 'the aforementioned' or 'the one we are talking about'. But see also the discussion of the Tok Pisin relative clauses §8.7.4.3.

[en] is a variant of [em], the third person pronoun, after [bilon] or [lon].

- (39) I gat pikinini insait long bel bilong en. Nau, em go askim ol man long kisim em long kanu na ol man tok: "Nogat, fulap ya, yu go painim narapela!"

i gat pikinini insait lon bel bilon en nau em go ask-im ol man
3.SBJ have child inside PREP belly GEN 3 now 3 go ask-TR PL person
lon kis-im em lon kanu na ol man tok nogat fulap ja ju go
PREP take-TR 3 PREP canoe and PL person say no full DEM 2 go
pain-im nara-pela
find-TR another-ADJ

'She had a child in her belly. Now, she went and asked everybody to take her in their canoe, and the people said, "No, we're full, go look for another one!"'

[lon] functions here as a complementizer, taking an entire clause as its complement.

- (40) Ol i toktok sem nau, em i go long narapela, ol tok sem. Em i go nau, las kanu nau, em i tok: "Nogat, mi fulap, mi gat pikinini bilong mi, na meri bilong mi yet. Yu go."

ol i toktok sem nau em i go lon nara-pela ol tok sem em i
3PL 3.SBJ say same now 3 3.SBJ go PREP other-ADJ PL say same 3 3.SBJ
go nau las kanu nau em i tok nogat mi fulap mi gat pikinini
go now last canoe now 3 3.SBJ say no 1 full 1 have child
bilon mi na meri bilon mi jet ju go
GEN 1 and woman GEN 1 self 2 go

'Since they spoke like that, she went to another, they said the same. So she went to the last canoe, and he said, "No, I'm full. I have my own children and my wife. You go."'

- (41) Ol man i tok, "Yu stap, yu gat bel, na nogat inap spes long yumi go."

ol man i tok ju stap ju gat bel na nogat inap spes
PL person 3.SBJ say 2 stay 2 have belly and NEG.EXIST enough space
lon jumi go
PREP 1PL.INCL go

'The people said, "You stay here, you're pregnant, and there is not enough space for you to come with us."'

- (42) Na ol i go painim nupela ailan na em i stap long olupela ples.

na ol i go pain-im nu-pela ailan na em i stap loŋ ol-pela ples
 and 3PL 3.SBJ go find-TR new-ADJ island and 3 3.SBJ stay PREP old-ADJ place
 ‘So they went to look for a new island, and she stayed in the old place.’

- (43) Nau, wanpela taim nau em bonim, em karim tupela pikinini boi, long insait long bel bilong en – olsem tupela pikinini, tupela boi, man.

nau wan-pela taim nau em bon-im em kar-im tu-pela pikinini boi loŋ insait
 now one-ADJ time now 3 bear-TR 3 carry-TR two-ADJ child boy PREP inside
 loŋ bel biloŋ en olsem tu-pela pikinini tu-pela boi man
 PREP belly GEN 3.OBL like two-ADJ child two-ADJ boy male
 ‘Now, one time she gave birth, she gave birth to two boy babies, she had been
 carrying two boy babies inside her belly – like, two babies, two boys, males.’

- (44) Tupela man nau, em i save wok hat long painim kaikai na abus bilong tupela.

tu-pela man nau em i save wok hat loŋ pain-im kaikai
 two-ADJ male now 3 3.SBJ HAB work hard PREP find-TR food
 na abus biloŋ tu-pela
 and meat GEN two-ADJ

‘Two males then, and she would work hard to find food and meat for the two of them.’

[abus] is a word meaning ‘bit of meat or game’ borrowed into Tok Pisin from Tolai.

- (45) Em i wokim olsem i go nau, tupela kamap bikpela yangpela man.

em i wok-im olsem i go nau tu-pela kamap bik-pela jaŋ-pela man
 3 3.SBJ do-TR thus 3.SBJ PROG now two-ADJ become big-ADJ young-ADJ man
 ‘She kept doing thus, and the two of them became big young men.’

[i wokim ... i go]. [i go] and [i kam], like [i stap], after the main verb add progressive aspect or direction to main verb.

- (46) Bikpela man nau, em tokim tupela stori long, pik ya, husat ya, em ol man, long ol man ronewe i go ya.

bik-pela man nau em tok-im tu-pela stori loŋ pik ja husat ja em ol
 big-ADJ man now 3 tell-TR two-ADJ story PREP pig DEM who DEM 3 PL
 man loŋ ol man ronewe i go ja
 person PREP PL person run_away 3.SBJ go DEM

‘Now that they were big men, she told the two of them the story of the pig who the people had run away from.’

- (47) Em stori long tupela pinis nau, em wokim supia na soim tupela long ol na sutim na ol sutim na yusim holim gut.

em stori loŋ tu-pela pinis nau em wok-im supia na so-im tu-pela
 DEM story PREP two-ADJ finish now 3 make-TR spear and show-TR two-ADJ
 loŋ ol na sut-im na ol sut-im na jus-im hol-im gut
 PREP 3PL and shoot-TR and 3PL shoot-TR and use-TR hold-TR well

‘Having finished the story for them, she made spears and showed them to them and hurled them, and they hurled and used them, handled them well.’

[soim tupela loŋ ol] = ‘And she showed them to the two.’ The Tok Pisin construction is actually very different from the English equivalent. In the Tok Pisin sentence, [tupela] ‘the two’ is the direct object of the verb, and [ol] ‘them’ has to be marked as the object of a preposition ([loŋ]). In English, however, *them* is the direct object, and *the two* has to be marked by the preposition *to*.

- (48) Nau bihain nau, em i go soim ples long we pik i save slip long en ya.

nau bihain nau em i go so-im ples loŋ we pik i save slip
 now later now 3 3.SBJ go show-TR place PREP where pig 3.SBJ HAB sleep
 loŋ en ja
 PREP 3.OBL DEM

‘Then later, she went and showed the place where the pig used to sleep.’

- (49) Ol i go antap long bikpela diwai na em i soim tupela long pik, draipela pik ya i save stap long en ya.

ol i go antap loŋ bik-pela diwai na em i so-im tu-pela loŋ
 3PL 3.SBJ go up PREP big-ADJ tree and 3 3.SBJ show-TR two-ADJ PREP
 pik drai-pela pik ja i save stap loŋ en ja
 pig huge-ADJ pig DEM 3.SBJ HAB stay PREP 3.OBL DEM

‘They went up a big tree, and she showed the two of them the pig, the huge pig that used to stay there.’

- (50) Nau ol i kam bek nau, wokim planti supia tru. Wokim planti supia nau, bihain ol i go bildim flatfom antap long diwai.

nau ol i kam bek nau wok-im planti supia tru wok-im planti supia
 now 3PL 3.SBJ come back now make-TR many spear very make-TR many spear

nau bihain ol i go bild-im flatfom antap loŋ diwai
 now afterwards 3PL 3.SBJ go build-TR platform up PREP tree

‘So they came back then and made a very large number of spears. Having made a lot of spears, they then went and built a platform up in the tree.’

[tru] is an intensifier modifying [planti]. Note that it does not come directly before or after the word it modifies as an ordinary adverb would.

- (51) Fom antap long diwai nau, narapela moning nau, ol i go antap long diwai ya.
 fom antap loŋ diwai nau nara-pela moniŋ nau
 platform up PREP tree now another-ADJ morning now
 ol i go antap loŋ diwai ja
 3PL 3.SBJ go up PREP tree DEM
 ‘Once the platform was in the tree, another morning then, they went up the tree.’
- (52) Kisim supia na ol i go antap long flatfom long diwai. Ol i lukim, pik i go na i kam bek, dring wara na, waswas long wara.
 kis-im supia na ol i go antap loŋ flatfom loŋ diwai ol i luk-im
 take-TR spear and 3PL 3.SBJ go up PREP platform PREP tree 3PL 3.SBJ look-TR
 pik i go na i kam bek driŋ wara na waswas loŋ wara
 pig 3.SBJ go and 3.SBJ come back drink water and wash PREP water
 ‘Taking the spears they went up on the platform in the tree. They watched the pig coming and going, drinking water and washing in the water.’
- (53) Na ol tripela man ya, pikinini na mama bilong tupela ya, i go antap long wanpela bikpela diwai.
 na ol tri-pela man ja pikinini na mama bilonŋ tu-pela ja i
 and 3PL three-ADJ person DEM child and mother GEN two-ADJ DEM 3.SBJ
 go antap loŋ wan-pela bik-pela diwai
 go up PREP one-ADJ big-ADJ tree
 ‘And these three people, the children and their mother, went up a big tree.’
- (54) Na ol tripela i stap nau, pik ya laik kam, dring wara na waswas. Ol tripela, i tromoi supia na ol samting long . . .
 na ol tri-pela i stap nau pik ja laik kam driŋ wara na waswas
 and 3PL three-ADJ 3.SBJ stay now pig DEM about_to come drink water and wash
 ol tri-pela i tromoi supia na ol samtiŋ loŋ
 3PL three-ADJ 3.SBJ throw spear and PL thing PREP
 ‘And the three of them stayed there, and the pig came to drink water and wash. The three threw their spears and things at [it].’
- (55) Em pik i win olem ya, em kalap antap, na flatfom daunbelo ya bruk.
 em pik i win olem ja em kalap antap na flatfom daunbelo ja bruk
 DEM pig 3.SBJ win thus DEM 3 climb up and platform bottom DEM break
 ‘The pig was winning. He climbed up, and the bottom platform broke.’
 [em pik]. According to some authorities on Tok Pisin, the third person pronoun before a coreferential noun acts as a definite article: ‘the pig.’

- (56) Nau tripela kalap i go long antap wanpela flatfom antap long diwai, strongpela. Nau ol tromoi spia i go na pik i sotwin na em i laik i dai nau.

nau tri-pela kalap i go loŋ antap wan-pela flatfom antap loŋ diwai
 now three-ADJ climb 3.SBJ go PREP top one-ADJ platform top PREP tree
 stroŋ-pela nau ol tromoi spia i go na pik i sotwin na em i
 strong-ADJ now 3PL throw spear 3.SBJ go and pig 3.SBJ winded and 3 3.SBJ
 laik i dai nau
 about_to 3.SBJ die now

‘And the three climbed up on a platform high up in the tree, a strong one. And they threw their spears at him and the pig was winded, and he was about to die right then.’

- (57) Na tupela boi ya, i kam daun long seken flatfom na sutim em long supia. Na em i dai nau.

na tu-pela boi ja i kam daun loŋ seken flatfom na sut-im
 and two-ADJ boy DEM 3.SBJ come down PREP second platform and shoot-TR
 em loŋ supia na em i dai nau
 3 PREP spear and 3 3.SBJ die now

‘And these two boys came down to the second platform and shot him with their spears. And he died then.’

[loŋ] is used here in an instrumental sense.

- (58) Ol tripela kam daun na sutim em na brukim het bilong en long ston.

ol tri-pela kam daun na sut-im em na bruk-im het bilonjen loŋ ston
 3PL three-ADJ come down and shoot-TR 3 and break-TR head GEN 3.OBL PREP stone

‘The three of them then came down and shot him and broke his head with a stone.’

- (59) Nau, ol i stap nau, mama bilong tupela tok: “Bai yu wokim wanem long pik ya?”

nau ol i stap nau mama bilonj tu-pela tok
 now 3PL 3.SBJ be now mother GEN two-ADJ say
 bai ju wok-im wanem loŋ pik ja
 FUT 2 do-TR what PREP pig DEM

‘So there they were, and their mother said to them, “What will you do with the pig?”’

- (60) Na, tupela tok: “Yu mas katim skin bilong en na tromoi long solwara.

na tu-pela tok ju mas kat-im skin bilonjen na tromoi loŋ solwara
 and two-ADJ say 2 must cut-TR skin GEN 3.OBL and throw PREP sea

‘And the two said, “You must cut off its skin and throw it in the sea.’

- (61) Em bai karim i go long narapela ailan na ol man bilong yumi bai i kam bek.”
 em bai kar-im i go loŋ nara-pela ailan na ol man biloŋ
 3 FUT carry-TR 3.SBJ go PREP another-ADJ island and PL person GEN
 jumi bai i kam bek
 1PL.INCL FUT 3.SBJ come back
 ‘It [the sea] will carry it off to the other island, and our people will come back.’”
- (62) Nau ol katim skin bilong pik ya na ol ol karim i go long haus. Karim i go long haus nau, ol tromoi long solwara na karim i go.
 nau ol kat-im skin biloŋ pik ja na ol ol kar-im i go loŋ haus
 now 3PL cut-TR skin GEN pig DEM and 3PL 3PL carry-TR 3.SBJ go PREP house
 kar-im i go loŋ haus nau ol tromoi loŋ solwara na kar-im i go
 carry-TR 3.SBJ go PREP house now 3PL throw PREP sea and carry-TR 3.SBJ go
 ‘So they cut the pig’s skin and carried it to the house, then they threw it into the ocean, and it carried it away.’
- (63) Na draipela win na, wara i save –, i go, olsem, si bruk. Long moningtaim, si wokim olsem nau, solwara karim dispela skin bilong pik ya i go daun long ailan, narapela ailan ya.
 na drai-pela win na wara i save i go olsem si bruk loŋ
 and huge-ADJ wind and water 3.SBJ HAB 3.SBJ go like sea break PREP
 moniŋ-taim si wok-im olsem nau solwara kar-im dis-pela skin
 morning-time sea do-TR thus now ocean carry-TR this-ADJ skin
 biloŋ pik ja i go daun loŋ ailan nara-pela ailan ja
 GEN pig DEM 3.SBJ go down PREP island other-ADJ island DEM
 ‘And there was a tremendous wind, and the water kept going, as if the sea were raging. In the morning, the sea had been like that; the sea had carried the pig’s skin off to the island, the other island.’
- (64) Nau, wanpela lapun man, em i laik go, em i laik go we? Em i laik go long toilet o, we, nau em lukim disfela samting, bikpela samting, skin bilong pik ya long solwara.
 nau wan-pela lapun man em i laik go em i laik go
 now one-ADJ old man 3 3.SBJ about_to go 3 3.SBJ about_to go
 we em i laik go loŋ toilet o we nau em luk-im dis-fela
 where 3 3.SBJ about_to go PREP toilet or where now 3 look-TR this-ADJ
 samtiŋ bik-pela samtiŋ skin biloŋ pik ja loŋ solwara
 thing big-ADJ thing skin GEN pig DEM PREP sea
 ‘Now, one old man was about to go . . . Where was he about to go? He was about to go to the toilet or something, and he saw this thing, this big thing, this pig skin, in the ocean.’

- (65) Na em singautim olgeta man long ples bihain kam luk, sanap na lukluk.
 na em siŋaut-im olgeta man loŋ ples bihain kam luk sanap
 and 3 call-TR all person PREP village afterward come look stand_up
 na luk-luk
 and look-look
 ‘And he called everybody from the village to come then and look, stand up and have a look.’
- (66) Nau, ol sampela man tok olsem: “Ol meri mas karim pikinini na stap insait long haus, no ken kam arasait, ol man tasol kam arasait!”
 nau ol sam-pela man tok olsem ol meri mas kar-im pikinini na stap insait
 now PL some-ADJ man say thus PL woman must carry-TR child and stay inside
 loŋ haus no ken kam arasait ol man tasol kam arasait
 PREP house NEG can come outside PL man only come outside
 ‘And some of the men said, “The women must carry the children and stay indoors, they mustn’t come outside, only the men can come outside!”’
- (67) Nau, ol man kam sanap lukluk nau, ol i lukim olsem skin bilong pik bifo save kaikai ol man ya.
 nau ol man kam sanap luk-luk nau ol i luk-im olsem skin bilong pik
 now PL man come stand_up look-look now 3PL 3.SBJ look-TR thus skin GEN pig
 bifo save kaikai ol man ja
 previously HAB eat PL person DEM
 ‘So the men came and stood and stared, they looked at the very skin of the pig that used to eat people.’
- (68) Nau, ol i go kisim dispela skin bilong pik ya na tromoi i go antap long wesan.
 nau ol i go kis-im dis-pela skin bilong pik ja na tromoi
 then 3PL 3.SBJ go take-TR this-ADJ skin GEN pig DEM and throw
 i go antap loŋ wesan
 3.SBJ go top PREP sand
 ‘Now, they went and took this pig skin and threw it on top of the sand.’
- (69) Ol tingting bek long meri ya i gat bel bifo ya. Nau tupela i tok: “Nating tumoro samting yumi mas i go bek long ples bilong mitupela ya na lukim husat kilim dispela pik na tromwe.”
 ol tiŋtiŋ bek loŋ meri ja i gat bel bifo ja nau tu-pela i
 3PL think back PREP woman DEM 3.SBJ have belly previously DEM now two-ADJ 3.SBJ
 tok natiŋ tumoro samtiŋ jumi mas i go bek loŋ ples
 say perhaps tomorrow approximately 1PL.INCL must 3.SBJ go back PREP village
 bilong mi-tu-pela ja na luk-im husat kil-im dis-pela pik na tromwe
 GEN 1-DU-ADJ DEM and see-TR who kill-TR this-ADJ pig and throw_away
 ‘They thought back to the woman who had been pregnant before. Then two said, “Maybe tomorrow or so we should go back to our own place and see who killed this pig and threw it away.”’

- (70) Nau long moningtaim nau, ol i kisim olgeta samting bilong ol i go long ples.
 nau loŋ moniŋ-taim nau ol i kis-im olgeta samtiŋ bilonŋ ol
 now PREP morning-time now 3PL 3.SBJ take-TR all thing GEN 3PL
 i go loŋ ples
 3.SBJ go PREP home
 ‘And in the morning they took all their belongings and went home.’
- (71) Ol i go nau, lukim, na tripela man ya, tupela man ya na mama bilong entupela, ol kam sanap long nambis na lukluk long ol.
 ol i go nau luk-im na tri-pela man ja tu-pela man ja na mama
 3PL 3.SBJ go now look-TR and three-ADJ people DEM two-ADJ man DEM and mother
 bilonŋ en-tu-pela ol kam sanap loŋ nambis na luk-luk loŋ ol
 GEN 3-DU-ADJ 3PL came stand_up PREP beach and look-look PREP 3PL
 ‘They went, looked, and the three people, the two men and their mother, they had come to stand on the beach and look at them.’
- (72) Ol i lukim kanu na olgeta samting. Ol man ya long kanu ol i lukluk i go na lukim tripela sanap.
 ol i luk-im kanu na olgeta samtiŋ ol man ja loŋ kanu ol i
 3PL 3.SBJ look-TR canoe and all thing PL person DEM PREP canoe 3PL 3.SBJ
 luk-luk i go na luk-im tri-pela sanap
 look-look 3.SBJ go and look-TR three-ADJ stand_up
 ‘They looked at the canoes and everything. The people in the canoes looked towards the three people standing there.’
- (73) Ol i wokim kaikai na redi i stap long ol man kam bek na kaikai.
 ol i wok-im kaikai na redi i stap loŋ ol man kam bek na kaikai
 3PL 3.SBJ make-TR food and ready 3.SBJ stay PREP PL people come back and eat
 ‘They had made food, and it was sitting there ready for all the people to come back and eat.’
- (74) Em ol i kam nau, ol i wokim kaikai. Oltripela wokim kaikai na olgeta i stap.
 em ol i kam nau ol i wok-im kaikai ol-tri-pela wok-im kaikai
 3 3PL 3.SBJ come now 3PL 3.SBJ make-TR food 3PL-three-ADJ make-TR food
 na olgeta i stap
 and all 3.SBJ be_there
 ‘They came and made food. The three of them made food, and there they all were.’

- (75) Ol man i kam kaikai nau, oltripela bai i go daun, ol i go, ol i go daun na i go, i go daunbilo tru.

ol man i kam kaikai nau ol-tri-pela bai i go daun ol i go
 3PL person 3.SBJ come eat now 3PL-three-ADJ FUT 3.SBJ go down 3PL 3.SBJ go

ol i go daun na i go i go daunbilo tru
 3PL 3.SBJ go down and 3.SBJ go 3.SBJ go deep very

‘After the people came and ate, the three of them went down into the sea. They went down, down, down, down to the very depths.’

[bai] usually acts as a future tense marker, but, according to Sankoff, here it is being used in an innovative way as a past punctual marker.

- (76) Bai oltripela i go insait long solwara na, pik ya ol kilim bipo ya bai kamap olsem draipela ston, blakpela ston.

bai ol-tri-pela i go insait loŋ solwara na pik ja ol kil-im bipo
 FUT 3PL-three-ADJ3.SBJ go inside PREP sea and pig DEM 3PL kill-TR earlier

ja bai kamap olsem drai-pela ston blak-pela ston
 DEM FUT become like huge-ADJ stone black-ADJ stone

‘Afterwards the three of them went down into the sea, and the pig that they had killed turned into an enormous stone, a black stone.’

- (77) Em tasol.

Em tasol

3 only

‘That’s all.’

8.8 Exercises

8.8.1 TOK PISIN PASSAGE

After reading the “Sketch of Tok Pisin” in this chapter (§8.7), try to translate the following Tok Pisin passage into English. For those words whose meaning you cannot guess, you may consult Mihalic (1971) or the online dictionary at <http://www.tok-pisin.com>. This passage is a fragment from a news story that appeared in the Tok Pisin newspaper magazine *Wantok* on September 3, 1975.

MOA BIA TAKIS

Mista Julius Chan, Minista bilong Fainens, i bin putim strongpela takis long bia na wiski na olkain siga na sigaret. Nau wanpela katon bia bai kostim wan kina moa; na olgeta peket sigaret bai kostim 8 toea moa.

Olsem tasol gavman i ting long winim 6 milien kina moa long yia.

Dispela yia Australia bai no givim mani inap long yia i go pinis. Na olsem gabman hia i mas painim sampela we o rot bilong winim mani. Na bikpela rot oltaim em kain kain takis.

Mista Chan i tok bia na sigaret i no samting bilong ol tumbuna o samting tru bilong pasin bilong PNG. Nogat.

Na Mista Somare i tok tu: Sapos man i no gat bia o sigaret, bai em i no hangre.

Na plenti meri tu bai i hepi long harim gavman i mekim hat liklik long ol man i lusim mani long bia. Planti i laik rausim bia, long wanem em i as bilong planti trabel long ples na long famili.

8.8.2 SOCIOLINGUISTIC ASPECTS OF CONTACT LANGUAGES

Although a number of contact languages such as Tok Pisin are widely used for serious purposes and even have some official status, many others have a low social status in their communities and only restricted uses in the public media.

For example, Hawaii Creole is seldom used in public except by local comedians. Thus, one gets the impression that the creole is fine to amuse people in public but is not appropriate for, say, political speeches or church sermons, even when the audiences consist entirely of Hawaii Creole speakers.

Finally, although most people would not even think of making fun of someone who speaks a foreign language like Spanish or Russian, many people react with great amusement to contact languages which are based on their own native languages. What, for example, was your own reaction, as a speaker or reader of Standard English, when you first encountered Tok Pisin?

Suggest some reasons that might explain (a) the low social status of many contact languages, (b) the restricted use of such languages for certain purposes such as entertainment, and (c) why many people react to contact languages with amusement.

8.8.3 HAWAII CREOLE

Examine the following excerpt from Ronald Nishihara's story "Lady's Man", which appeared in the May 1979 edition of the *Ad Hoc* magazine published by Iolani School, Honolulu, Hawaii. It attempts to write down a variety of Hawaii Creole that contains some elements of Standard American English.

Is there anything in this fragment to indicate that what we are dealing with here is a creole and not simply another dialect of American English? If so, point out the typical creole features and discuss them briefly. If not, point out some of the major features that differentiate Hawaii Creole from Standard American English.

Afta we come out of *Sleeping Beauty*, we went someplace for grind. Ho man! Da snack bar at Kaimuki Bowling alley got good kine food.

Den we went go around in da small parking lot doing about thirty. Ho man, she got turn white one time when I almost went hit one other car.

Afta, she started saying something about one headache, so I go toss her one aspirin I found on da floor of da car and said “Take dis den you gon’ feel mo betta.” Den she went go say “No, I really don’t feel very well, I think I should just go home and rest a little while.”

So I went go take her home at about one twenty on da straightways. Den she go turn white one time when I went go make one U-turn in her neighbor’s front lawn and pulled up right in front of her house.

... As I was driving home, I told myself what I told myself when I was twelve years old, I said “Eh, I ste one lady’s man.”

8.9 Suggested readings

8.9.1 GENERAL

- ✘ *The Cambridge handbook of endangered languages* (Austin & Sallabank 2015). 23 chapters cover the extent of language endangerment and measures to document and revitalize them. Of special interest: “Introduction” by the editors, and “Revitalization of endangered languages” by Leanne Hinton.
- ✘ *Contact languages: A comprehensive guide* (Bakker & Matras 2013).
- ✘ *Contact languages: A wider perspective* (Thomason 1997). Detailed sketches of 12 contact languages that did not arise directly from contact with Europeans.
- ✘ *The emergence of pidgin and creole languages* (Siegel 2008). General introduction with case studies from Pidgin Fijian, Melanesian Pidgin, Hawaii Creole, New Caledonian Tayo, and Australian Kriol.
- ✘ *Endangered languages: Critical concepts in linguistics* (Austin & McGill 2011). Volume 2 begins with an overview of language documentation.
- ✘ *The green book of language revitalization in practice* (Hinton & Hale 2001). A seminal collection of articles describing language revitalization efforts in various parts of the world, with special attention to the United States. Of special interest: “Language revitalization: An overview” by Leanne Hinton and “Diversity in local language maintenance and restoration: A reason for optimism” by Anna Ash, Jessie Little Doe Fermino, and Ken Hale.

- ❑ *The handbook of pidgin and creole studies* (Kouwenberg & Singler 2009). A state-of-the-art reference volume with 26 chapters by international experts.
- ❑ *An introduction to pidgins and creoles* (Holm 2000). This introduction to pidgins and creoles includes theories of how they develop and case studies of seven creoles from around the world.
- ❑ *Language contact, creolization, and genetic linguistics* (Thomason & Kaufman 1988). The authors subsume creolization under the genetic model of language development. Included are sketches of seven creoles and a detailed treatment of the development of dialects of English and its near cousins.
- ❑ *The making of a mixed language: The case of Maà/Mbugu* (Mous 2003).
- ❑ *The mixed language debate: Theoretical and empirical advances* (Matras & Bakker 2013).
- ❑ *Pidgins and creoles* (Mufwene 2001). Overview of the meanings of the basic terminology, and a cogent presentation of arguments casting doubt on traditional views of the origins of pidgins and creoles.
- ❑ *Pidgins and creoles* (Singh 2000). A very basic introduction to pidgins and creoles and theories of how they develop.
- ❑ *Pidgins and creoles: An introduction* (Arends et al. 1994). Though billed as an introduction, this book describes pidgins and creoles and theories of their origin in considerable detail. Included are eight 5- to 10-page sketches of specific languages.

8.9.2 TOK PISIN

- ❑ *A new course in Tok Pisin (New Guinea Pidgin)* (Dutton & Thomas 1985). After many years, still the standard introductory textbook for this language.
- ❑ *Towards a reference grammar of Tok Pisin: An experiment in corpus linguistics* (Verhaar 1995). (A descriptive grammar based on a large computerized corpus.)

8.9.3 SIGN LANGUAGES

- ❑ *Sign language: An international handbook* (Pfau et al. 2012). An encyclopedic reference work of 44 chapters touching on all facets of sign language, including typology, sign language grammar, documentation and transcription, and psychological, social, and neural aspects of sign language learning and use.
- ❑ *Sign language and linguistic universals* (Sandler & Lillo-Martin 2006). A basic but highly inclusive introduction to ASL, paying special attention to similarities and differences between signed and spoken language.

- ✘ *Sign language structure: An outline of the visual communication systems of the American deaf* (Stokoe 1960). Mainly of historical interest now, it was the first published analysis of ASL attempting to follow the linguistic procedures of its time.
- ✘ *Sign languages* (Brentari 2010). Collection of articles on sign languages from around the world, dealing with their history and linguistic structure and with national policies regarding sign languages.
- ✘ *The signs of language* (Klima & Bellugi 1979). A classic early collection of articles by linguists and sign language specialists covering the state of the art at the time of publication, with a focus on the syntax and phonology of ASL.

REFERENCES

- Abondolo, Daniel (ed.). 1998. *The Uralic languages*. London: Routledge. (Cited 3.8.3)
- Adelaar, W. F. H. 1977. *Tarma Quechua: Grammar, texts, dictionary*. Lisse: Peter de Ridder Press. (Cited 7.29.5)
- Ager, Simon. 2015. Omniglot: The online encyclopedia of writing systems and languages. (<http://www.omniglot.com>) (Cited 2.2.2)
- Anderson, Gregory D. S. (ed.). 2008. *The Munda languages*. London: Routledge. (Cited 4.13.7)
- Arends, Jacques & Muysken, Pieter & Smith, Norval (eds.). 1994. *Pidgins and creoles: An introduction*. Amsterdam: Benjamins. (<https://doi.org/10.1075/cll.15>) (Cited 8.9.1)
- Arnott, David W. 1964. Downstep in the Tiv verbal system. *African Language Studies* 5. 34–51. (Cited 5.3.7)
- Ash, Anna & Fermino, Jessie Little Doe & Hale, Ken. 2001. Diversity in local language maintenance and restoration: A reason for optimism. In Hinton, Leanne & Hale, Ken (eds.), *The green book of language revitalization in practice*, 19–35. San Diego: Academic Press. (https://doi.org/10.1163/9789004261723_003) (Cited 8.6.4)
- Austerlitz, Robert. 1966. *Finnish reader and glossary*. Bloomington: Indiana University Press. (Cited 3.6.5)
- Austin, Peter K. (ed.). 2008. *One thousand languages: Living, endangered, and lost*. Berkeley: University of California Press. (Cited 1.4.3)
- Austin, Peter K. & McGill, Stuart (eds.). 2011. *Endangered languages*. (4 vols.) London: Routledge. (Cited 8.9.1)
- Austin, Peter K. & Sallabank, Julia (eds.). 2015. *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press. (Cited 8.9.1)
- Australian Bureau of Statistics. 2013. (<http://www.abs.gov.au>) (Cited 6.3)
- Australian Institute of Aboriginal and Torres Strait Islander Studies. 2005. *National Indigenous Languages Survey report 2005*. Canberra: Department of Communications, Information Technology and the Arts. (Cited 6.3)
- Bagemihl, Bruce. 1998. Maximality in Bella Coola (Nuxalk). In Czaykowska-Higgins, Ewa & Kinkade, M. Dale (eds.), *Salish languages and linguistics: Theoretical and descriptive perspectives*, 71–98. Berlin: Mouton de Gruyter. (Cited 7.13)
- Bakker, Peter & Matras, Yaron (eds.). 2013. *Contact languages: A comprehensive guide*. Boston: de Gruyter. (Cited 8.9.1)
- Ball, Martin & Müller, Nicole (eds.). 2009. *The Celtic languages*. 2nd edn. New York: Routledge. (Cited 3.8.2)
- Bellwood, Peter. 1991. The Austronesian dispersal and the origin of languages. *Scientific American* 265(1). 88–93. (Cited 6.1.3, 6.7.2)
- Bender, M. Lionel. 1996. *The Nilo-Saharan languages: A comparative essay*. Munich: Lincom Europa. (Cited 5.9.3)

- Bendor-Samuel, John & Hartell, Rhonda L. (eds.). 1989. *The Niger-Congo languages: A classification and description of Africa's largest language family*. Lanham: University Press of America. (Cited 5.9.4)
- Benedict, Paul K. 1942. Thai, Kadai, and Indonesian: A new alignment in Southeastern Asia. *American Anthropologist* 44(4). 576–601. (<https://doi.org/10.1525/aa.1942.44.4.02a00040>) (Cited 4.5)
- Bengtson, John D. & Ruhlen, Merritt. 1994. Global etymologies. In Ruhlen, Merritt (ed.), *On the origin of languages: Studies in linguistic taxonomy*, 277–336. Stanford: Stanford University Press. (Cited 1.1.3)
- Blake, Barry J. 1991. *Australian aboriginal languages: A general introduction*. 2nd edn. St. Lucia: University of Queensland Press. (Cited 6.7.4)
- Blench, Roger. 2000. The classification of Nilo-Saharan. *Afrika und Übersee* 83. 293–307. (Cited 5.2)
- Blench, Roger. 2006. *Archaeology, language, and the African past*. Lanham: AltaMira. (Cited 5.9.1)
- Bliss, Heather Anne. 2013. *The Blackfoot configurationality conspiracy: Parallels and differences in clausal and nominal structures*. Vancouver: University of British Columbia. (Doctoral dissertation.) (Cited 7.3)
- Bloomfield, Leonard. 1925. On the sound-system of Central Algonquian. *Language* 1(4). 130–156. (<https://doi.org/10.2307/409540>) (Cited 7.3)
- Blust, Robert. 1985. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26(1). 45–67. (Cited 6.1.3)
- Blust, Robert. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Zeitoun, Elizabeth & Li, Paul Jen-kuei (eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics*, 31–94. Taipei: Academia Sinica. (Cited 6.1.1)
- Blust, Robert. 2009. *The Austronesian languages*. Canberra: Pacific Linguistics, Australian National University. (Cited 6.7.2)
- Bondarko, Liya V. 2005. Phonetic and phonological aspects of the opposition of ‘soft’ and ‘hard’ consonants in the modern Russian language. *Speech Communication* 47. 7–14. (<https://doi.org/10.1016/j.specom.2005.03.012>) (Cited 3.5.3)
- Boone, Elizabeth Hill. 2004. Beyond writing. In Houston, Stephen D. (ed.), *The first writing: Script invention as history and process*, 313–348. Cambridge: Cambridge University Press. (Cited 2.2.1.4)
- Bowern, Claire & Koch, Harold (eds.). 2004. *Australian languages: Classification and the comparative method*. Amsterdam: Benjamins. (<https://doi.org/10.1075/cilt.249>) (Cited 6.7.4)
- Breinstrup, Thomas (ed.). 2013. Le angulo linguistic: Ascoltar musica (MP3) [The Linguistic Corner: Listening to music (MP3)]. (<http://www.interlingua.com/linguistic/mp3>) (Cited 8.2)
- Brentari, Diane (ed.). 2010. *Sign languages*. New York: Cambridge University Press. (Cited 8.9.3)
- Brown, Andrew. 2003, June 26. A way with words. *The Guardian*. (<http://www.theguardian.com/science/2003/jun/26/scienceinterviews.artsandhumanities>) (Cited 3.4)
- Brown, Keith (ed.). 2006. *Encyclopedia of language and linguistics*. 2nd edn. Amsterdam: Elsevier. (Cited 1.4.3)
- Campbell, George L., & King, Gareth. 2011. *The Routledge concise compendium of the world's languages*. 2nd edn. New York: Routledge. (Cited 1.4.3)

- Campbell, Lyle. 1979. Middle American languages. In Campbell, Lyle & Mithun, Marianne, *The languages of Native America: Historical and comparative assessment*, 902–1000. Austin: University of Texas Press. (Cited 7.29.4)
- Campbell, Lyle. 1988. Review of *Language in the Americas*, by Joseph H. Greenberg. *Language* 64(3). 591–615. (<https://doi.org/10.2307/414535>) (Cited 7.13, 7.29.1)
- Campbell, Lyle. 1997. *American Indian languages: The historical linguistics of Native America*. New York: Oxford University Press. (Cited Chapter 7, 7.29.1)
- Campbell, Lyle. 2011. [Review of] *The Dene-Yeniseian connection*, edited by James Kari and Ben A. Potter. *International Journal of American Linguistics* 77(3). 445–451. (<https://doi.org/10.1086/660977>) (Cited 4.2)
- Campbell, Lyle. 2013. *Historical linguistics: An introduction*. 3rd edn. Cambridge: MIT Press. (Cited 1.4.4)
- Campbell, Lyle & Kaufman, Terrence. 1976. A linguistic look at the Olmecs. *American Antiquity* 41(1). 80–89. (<https://doi.org/10.2307/279044>) (Cited 7.16)
- Campbell, Lyle & Kaufman, Terrence. 1985. Mayan linguistics: Where are we now? *Annual Review of Anthropology* 14. 187–198. (<https://doi.org/10.1146/annurev.an.14.100185.001155>) (Cited 7.17)
- Campbell, Lyle, & Poser, William J. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press. (Cited 1.4.4)
- Cǎo, Cuiyún. 1987. Miáoyáyǔ tèdiǎn gǎiyào (苗瑶语特点概要) [Outline of the characteristics of Miáoyáo]. In Zhōngyāng Mǐnzhú Xuéyuàn. Shǎoshǔ Mǐnzhú Yǔyán Yánjiùsuǒ, Zhōngguó shǎoshǔ mǐnzhú yǔyán, 403–405. Chengdu: Sìchuān Mǐnzhú Chūbǎnshè. (Cited 4.4)
- Capell, A. 1969. *A survey of New Guinea languages*. Sydney: Sydney University Press. (Cited Tab. 6.14)
- Cardona, George & Jain, Danesh (eds.). 2007. *The Indo-Aryan languages*. New York: Routledge. (Cited 3.8.2)
- Chacon, Thiago Costa. 2012. *The phonology and morphology of Kubeo: The documentation, theory, and description of an Amazonian language*. Mānoa: University of Hawai‘i at Mānoa. (Doctoral dissertation.) (Cited 7.19)
- Chadwick, John. 1990. *The decipherment of Linear B*. 2nd edn. Cambridge: Cambridge University Press. (Cited 3.1.4)
- Chang, Charles B. 2006. Tense consonants in Korean revisited: A crosslinguistic perceptual study. In Chang, Charles B. et al. (eds.), *CamLing 2006: Proceedings of the Fourth University of Cambridge Postgraduate Conference in Language Research*, 35–42. Cambridge: Cambridge Institute of Language Research. (Cited 4.1.4)
- Chao, Yuen Ren. 1930. ə sistim əv “toun letəz” (A system of “tone letters”). *Lə mɛ:trə fɒnetik (Le Maître Phonétique)* 1930(30). 24–27. (Cited 4.10.2.3)
- Chao, Yuen Ren. 1957. *Mandarin primer: An intensive course in spoken Chinese*. Cambridge: Harvard University Press. (Cited 4.10.4)
- Childs, G. Tucker. 2003. *An introduction to African languages*. Amsterdam: Benjamins. (<https://doi.org/10.1075/z.121>) (Cited 5.9.1)
- China, People’s Republic. 2000. Law of the People’s Republic of China on the standard spoken and written Chinese language. (The National People’s Congress of the People’s Republic of China web site, http://www.npc.gov.cn/englishnpc/Law/2007-12/11/content_1383540.htm) (Cited 4.10.1)

- Christian Brothers. 1993. *New Irish grammar*. Dublin: C. J. Fallon. (Cited 3.1.1)
- Clement, Victoria. 2008. Emblems of independence: Script choice in post-Soviet Turkmenistan. *International Journal of the Sociology of Language* 192. 171–185. (<https://doi.org/10.1515/IJSL.2008.042>) (Cited 4.1.1)
- Colapinto, John. 2007, April 16. The interpreter: Has a remote Amazonian tribe upended our understanding of language? *The New Yorker*. (<http://www.newyorker.com/magazine/2007/04/16/the-interpreter-2>) (Cited 7.24)
- Collinder, Björn. 1965. *An introduction to the Uralic languages*. Berkeley: University of California Press. (Cited 3.2.7)
- Comrie, Bernard. 1981. Paleosiberian and other languages. In Comrie, Bernard, *The languages of the Soviet Union*, 238–278. Cambridge: Cambridge University Press. (Cited 4.2, 4.13.3)
- Comrie, Bernard (ed.). 1987. *The world's major languages*. New York: Oxford University Press. (Cited 1.1)
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. 2nd edn. Chicago: University of Chicago Press. (Cited 1.4.6)
- Comrie, Bernard. 2001. Languages of the world. In Aronoff, Mark & Rees-Miller, Janie (eds.), *The handbook of linguistics*, 19–42. Malden: Blackwell. (Cited 1.4.2)
- Comrie, Bernard (ed.). 2009. *The world's major languages*. 2nd edn. New York: Routledge. (Cited 1.4.3, 3.8.1)
- Comrie, Bernard & Matthews, Stephen & Polinsky, Maria. 2003. *The atlas of languages: The origin and development of languages throughout the world*. Rev. edn. London: Quarto. (Cited 1.4.3)
- Comrie, Bernard & Polinsky, Maria. 1998. The great Daghestanian case hoax. In Siewierska, Anna & Song, Jae Jung (eds.), *Case, typology and grammar: In honor of Barry J. Blake*, 95–114. Amsterdam: Benjamins. (<https://doi.org/10.1075/tsl.38>) (Cited 3.3.2)
- Cook, Eung Do & Gerds, Donna B. (eds.). 1984. *The syntax of Native American languages*. Orlando: Academic Press. (Cited 7.29.1)
- Cook, John. 2001, November 18. Symbol making. *New York Times Magazine*. (<http://www.nytimes.com/2001/11/18/magazine/18PROCESS.html>) (Cited 2.1)
- Cooper, Jerrold S. 2004. Babylonian beginnings: The origin of the cuneiform writing system in comparative perspective. In Houston, Stephen D. (ed.), *The first writing: Script invention as history and process*, 71–99. Cambridge: Cambridge University Press. (Cited 2.2.1.4)
- Coulmas, Florian (ed.). 1996. *The Blackwell encyclopedia of writing systems*. Oxford: Blackwell. (Cited 2.4)
- Croft, William. 2002. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press. (Cited 1.4.6)
- Crowley, Terry, & Dixon, R. M. W. 1981. Tasmanian. In Dixon, R. M. W. & Blake, Barry J. (eds.), *Handbook of Australian Languages* vol. 2, 395–421. Amsterdam: Benjamins. (<https://doi.org/10.1075/z.hal2>) (Cited 8.6)
- Crystal, David. 2003. *English as a global language*. 2nd edn. Cambridge: Cambridge University Press. (Cited 3.1.2)
- Cubberley, Paul. 2002. *Russian: A linguistic introduction*. Cambridge: Cambridge University Press. (Cited 3.8.6)
- Daniels, Peter T. 1990. Fundamentals of grammatology. *Journal of the American Oriental Society* 110(4). 727–731. (Cited 2.1.3.4, 2.4)

- Daniels, Peter T. 1996. The study of writing systems. In Daniels, Peter T. & Bright, William (eds.), *The world's writing systems*, 3–17. New York: Oxford University Press. (Cited 2.1.3.3)
- Daniels, Peter T. & Bright, William (eds.). 1996. *The world's writing systems*. New York: Oxford University Press. (Cited 2.2.2, 2.4)
- Davis, Jeffrey E. 2010. *Hand talk: Sign language among American Indian nations*. Cambridge: Cambridge University Press. (Cited 8.5.2)
- DeFrancis, John. 1984. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawai'i Press. (Cited 2.4)
- DeFrancis, John. 1989. *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawai'i Press. (Cited 2.4)
- Demolin, Didier. 2002. The search for primitives in phonology and the explanation of sound patterns: The contribution of fieldwork studies. In Gussenhoven, Carlos & Warner, Natasha (eds.), *Laboratory phonology* 7.455–514. Berlin: de Gruyter. (Cited 5.2.2)
- Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North Holland Publishing Co. (Cited 7.28.1)
- Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Dallas: Summer Institute of Linguistics. (Cited 7.25, 7.28.1)
- Derbyshire, Desmond C. & Pullum, Geoffrey K. 1981. Object-initial languages. *International Journal of American Linguistics* 47(3). 192–214. (Cited 7.25)
- Derbyshire, Desmond C. & Pullum, Geoffrey K. (eds.). 1986–1998. *Handbook of Amazonian languages*. New York: de Gruyter. (Cited 7.29.5)
- de Rijk, Rudolf P. G. 2007. *Standard Basque: A progressive grammar*. Cambridge: MIT Press. (Cited 3.8.5)
- Diller, Anthony & Edmondson, Jerry & Luo, Yongxian (eds.). 2011. *The Tai-Kadai languages*. New York: Routledge. (Cited 4.13.6)
- Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42(2). 214–261. (Cited 5.2.5)
- Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press. (Cited 6.5.1, 6.5.4, 6.5.6.1, 6.5.7, 6.7.5)
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press. (<https://doi.org/10.1017/CBO9780511611896>) (Cited 6.5.1)
- Dixon, R. M. W. 2011. *The languages of Australia*. Cambridge: Cambridge University Press. (Reissue of the 1980 edition.) (Cited 6.5.1, 6.7.4)
- Dixon, R. M. W. & Aikhenvald, Alexandra Y. (eds.). 1999. *The Amazonian languages*. Cambridge: Cambridge University Press. (Cited 7.29.5)
- Dixon, R. M. W. & Blake, Barry J. (eds.). 1979–2000. *Handbook of Australian languages*. Amsterdam: Benjamins; New York: Oxford University Press. (Cited 6.7.4)
- Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. The world atlas of linguistic structures online. Munich: Max Planck Institute for Evolutionary Anthropology. (<http://wals.info/>) (Cited 1.4.1)
- Dutton, Tom & Thomas, Dicks. 1985. *A new course in Tok Pisin (New Guinea Pidgin)*. Canberra: Department of Linguistics, Australian National University. (Cited 8.9.2)
- Egerod, Søren. 1991. Far Eastern languages. In Lamb, Sydney M. & Mitchell, Douglas E. (eds.), *Sprung from some common source: Investigations into the prehistory of languages*, 205–231. Stanford: Stanford University Press. (Cited 4.13.1)

- Elbert, Samuel & Pukui, Mary Kawena. 1979. *Hawaiian grammar*. Honolulu: University of Hawai'i Press. (Cited 6.4)
- Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46(4). 621–646. (<https://doi.org/10.1086/431525>) (Cited 7.24)
- Fagundes, Nelson J. R. & Kanitz, Ricardo & Eckert, Roberta & Valls, Ana C. S. & Bogo, Mauricio R. & Salzano, Francisco M. & Smith, David Glenn, et al. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *The American Journal of Human Genetics* 82(3). 583–592. (<https://doi.org/10.1016/j.ajhg.2007.11.013>) (Cited Chapter 7)
- Ferguson, Charles A. 1963. Assumptions about nasals: A sample study in phonological universals. In Greenberg, Joseph H. (ed.), *Universals of language*, 50–60. Cambridge: MIT Press. (Cited 1.2.3)
- Finer, Daniel L. 1985. The syntax of switch-reference. *Linguistic Inquiry* 16(1). 35–55. (Cited 7.8)
- Foley, William A. 1986. *The Papuan languages of New Guinea*. New York: Cambridge University Press. (Cited 6.7.3)
- Foley, William A. 1992. New Guinea languages. In Bright, William et al. (eds.), *International encyclopedia of linguistics*, vol. 3, 86–91. New York: Oxford University Press. (Cited 6.2)
- Fonseca, Felicia. 2015. Navajo Nation loosens language requirements for top leaders. (<http://bigstory.ap.org/article/of2bac645cd34a71835782eb377274c3/election-focuses-navajo-language-fluency-requirement>) (Cited 8.6.2)
- Fortson, Benjamin W., IV. 2010. *Indo-European language and culture: An introduction*. 2nd edn. Oxford: Wiley-Blackwell. (Cited 3.8.2)
- Gelb, I. J. 1963. *A study of writing*. 2nd, rev. edn. Chicago: University of Chicago Press. (Cited 2.1.3.3, 2.2.2)
- Gilsdorf, Thomas E. 2012. *Introduction to cultural mathematics: With case studies in the Otomies and Incas*. Hoboken: Wiley. (Cited 7.14)
- Goddard, Cliff. 2005. *The languages of East and Southeast Asia: An introduction*. New York: Oxford University Press. (Cited 4.13.1)
- Goebel, Ted & Waters, Michael R. & O'Rourke, Dennis H. 2008. The Late Pleistocene dispersal of modern humans in the Americas. *Science* 319. 1497–1502. (<https://doi.org/10.1126/science.1153569>) (Cited Chapter 7)
- Goldin-Meadow, Susan. 2005. *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language*. New York: Psychology Press. (Cited 8.5.2)
- Goldwasser, Orly. 2010. How the alphabet was born from hieroglyphs. *Biblical Archaeology Review* 36(2). 38–51. (Cited 2.2.1.3)
- Government of India. Office of the Registrar General & Census Commissioner. 2006. Abstract of speakers' strength of languages and mother tongues – 2001. (http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.aspx) (Accessed 2016-05-26.) (Cited 3.1.8)
- Greenberg, Joseph H. 1954. A quantitative approach to the morphological typology of language. In Spencer, Robert F. (ed.), *Method and perspective in anthropology*, 192–220. Minneapolis: University of Minnesota Press. (Cited 1.2.1)
- Greenberg, Joseph H. 1963. *The languages of Africa*. Bloomington: Indiana University Press. (Cited 1.1.3, Chapter 5, 5.2, 5.4)

- Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph (ed.), *Universals of language*, 73–113. 2nd edn. Cambridge: MIT Press. (Cited 1.2.2)
- Greenberg, Joseph H. 1971. The Indo-Pacific hypothesis. In Sebeok, Thomas A. (ed.), *Current trends in linguistics*, vol. 8, 807–871. The Hague: Mouton. (Cited 1.1.3, Chapter 6, 6.2–3)
- Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press. (Cited 1.1.3, 7.13, Chapter 7, 7.29.1)
- Gregersen, Edgar A. 1977. *Language in Africa: An introductory survey*. New York: Gordon and Breach. (Cited 5.3.3)
- Güldemann, Tom & Vossen, Rainer. 2000. Khoisan. In Heine, Bernd & Nurse, Derek (eds.), *African languages: An introduction*, 99–122. Cambridge: Cambridge University Press. (Cited 5.9.5)
- Gustason, Gerilee & Zawolkow, Esther. 1993. *Signing Exact English*. Los Alamitos: Modern Signs Press. (Cited 8.5.1)
- Haas, Mary R. 1951. The Proto-Gulf word for *water* (with notes on Siouan-Yuchi). *International Journal of American Linguistics* 17(2). 71–79. (Cited 7.4)
- Haas, William. 1976. Writing: The basic options. In Haas, William (ed.), *Writing without letters*, 131–208. Manchester: Manchester University Press. (Cited 2.1.1)
- Hale, Ken. 1981. *On the position of Walbiri in a typology of the base*. Bloomington: Indiana University Linguistics Club. (Cited 6.3.3)
- Hale, Ken & Nash, David. 1997. Damin and Lardil phonotactics. In Tryon, Darrell & Walsh, Michael (eds.), *Boundary rider: Essays in honour of Geoffrey O'Grady*, 247–259. Canberra: Pacific Linguistics, Australian National University. (Cited 8.2)
- Hall, Robert A., Jr. 1944. Chinese Pidgin English grammar and texts. *Journal of the American Oriental Society* 64(3). 95–113. (<https://doi.org/10.2307/594237>) (Cited 8.3)
- Hammarström, Harald & Forkel, Robert & Haspelmath, Martin & Bank, Sebastian. 2016. Glottolog 2.7. Jena: Max Planck Institute for the Science of Human History. (<http://glottolog.org>) (Cited 1.4.1)
- Harbert, Wayne. 2007. *The Germanic languages*. Cambridge: Cambridge University Press. (Cited 3.8.2)
- Harris, Martin & Vincent, Nigel (eds.). 1990. *The Romance languages*. Oxford: Oxford University Press. (Cited 3.8.2)
- Haspelmath, Martin & König, Ekkehard & Oesterreicher, Wulf & Raible, Wolfgang (eds.). 2001. *Language typology and language universals: An international handbook*. Berlin: de Gruyter. (Cited 1.4.6)
- Hauser, Marc D. & Chomsky, Noam & Fitch, W. Tecumseh. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569–1579. (<https://doi.org/10.1126/science.298.5598.1569>) (Cited 7.24)
- Haviland, John B. 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3(1). 3–45. (<https://doi.org/10.1525/jlin.1993.3.1.3>) (Cited 6.3.3)
- Heine, Bernd & Kuteva, Tania. 2006. *The changing languages of Europe*. New York: Oxford University Press. (Cited 8.1)
- Heine, Bernd & Nurse, Derek (eds.). 2000. *African languages: An introduction*. Cambridge: Cambridge University Press. (Cited 5.9.1)

- Heine, Bernd & Nurse, Derek (eds.). 2008. *A linguistic geography of Africa*. Cambridge: Cambridge University Press. (Cited 5.9.1)
- Hewitt, B. G. 1981. Caucasian languages. In Comrie, Bernard (ed.), *The languages of the Soviet Union*, 196–237. New York: Cambridge University Press. (Cited 3.8.4)
- Hewitt, B. G. 1987. Georgian: Ergative or active? *Lingua* 71. 319–340. ([https://doi.org/10.1016/0024-3841\(87\)90077-5](https://doi.org/10.1016/0024-3841(87)90077-5)) (Cited 3.3.3)
- Hinton, Leanne. 2001. Language revitalization: An overview. In Hinton, Leanne & Hale, Ken (eds.), *The green book of language revitalization in practice*, 3–18. San Diego: Academic Press. (Cited 8.6)
- Hinton, Leanne & Hale, Ken (eds.). 2001. *The green book of language revitalization in practice*. San Diego: Academic Press. (Cited 8.9.1)
- Hock, Hans Heinrich. 1991. *Principles of historical linguistics*. 2nd edn. Berlin: de Gruyter. (Cited 1.4.4)
- Hock, Hans Heinrich & Joseph, Brian D. 2009. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. 2nd edn. Berlin: de Gruyter. (Cited 1.4.4)
- Holes, Clive. 2004. *Modern Arabic: Structures, functions, and varieties*. Rev. edn. Washington: Georgetown University Press. (Cited 5.9.6)
- Holm, John. 2000. *An introduction to pidgins and creoles*. New York: Cambridge University Press. (Cited 8.9.1)
- Hopkins, Alberta Pualani. 1992. *Ka lei ha'aheo: Beginning Hawaiian*. Honolulu: University of Hawai'i Press. (Cited 6.4.4)
- Horne, Kibbey M. 1966. *Language typology: 19th and 20th century views*. Washington: Georgetown University Press. (Cited 1.2)
- Houston, Stephen D. (ed.). 2004. *The first writing: Script invention as history and process*. Cambridge: Cambridge University Press. (Cited 2.4)
- Hualde, José Ignacio & Ortiz de Urbina, Jon. 2003. *A grammar of Basque*. Berlin: de Gruyter. (Cited 3.4)
- Hunter, William C. 1938. *The "fan kwae" at Canton before treaty days, 1825–1844*. Shanghai: Oriental Affairs. (Cited 8.3)
- Hyman, Larry M. 2009. How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* 31. 213–238. (<https://doi.org/10.1016/j.langsci.2008.12.007>) (Cited 1.2.3)
- Jacobsen, William H., Jr. 1967. Switch-reference in Hokan-Coahuiltecan. In Hymes, Dell H. & Bittle, William E. (eds.), *Studies in Southwestern ethnolinguistics*, 238–263. The Hague: Mouton. (Cited 7.8)
- Jacobson, Steven A. & Jacobson, Anna W. 1995. *A practical grammar of the Central Alaskan Yup'ik Eskimo language*. Fairbanks: Alaska Native Language Center and Program, University of Alaska. (Cited 7.29.3)
- Jacques, Guillaume. 2012. A new transcription system for Old and Classical Tibetan. *Linguistics of the Tibeto-Burman Area* 35(2). 89–96. (Cited 4.11.2)
- Jacobson, Roman. 1958. Typological studies and their contribution to historical comparative linguistics. In Sivertsen, Eva (ed.), *Proceedings of the 8th International Congress of Linguists, Oslo, 1957*, 17–25. Oslo: Oslo University Press. (Cited 1.2.2)
- Janhunen, Juha. 1992. Uralic languages. In Bright, William (ed.), *International encyclopedia of linguistics*, vol. 4, 205–209. New York: Oxford University Press. (Cited 3.2.7)

- Janhunen, Juha (ed.). 2003. *The Mongolic languages*. New York: Routledge. (Cited 4.13.2)
- Johanson, Lars & Csató, Éva Á. (eds.). 2006. *The Turkic languages*. New York: Routledge. (Cited 4.13.2)
- Johnson, Ian. 2009. In China, the forgotten Manchu seek to rekindle their glory. *The Wall Street Journal*. (<http://online.wsj.com/news/articles/SB125452110732160485>) (Cited 4.1.3)
- Johnson, Keith. 2008. *Quantitative methods in linguistics*. Malden: Blackwell. (Cited 1.4.5)
- Kaiser, M. & Shevoroshkin, V. 1988. Nostratic. *Annual Review of Anthropology* 17. 309–329. (<https://doi.org/10.1146/annurev.an.17.100188.001521>) (Cited 1.4.4)
- Karimata, Shigehisa. 2015. Ryukyuan languages: A grammar overview. In Heinrich, Patrick & Miyara, Shinsho & Shimoji, Michinori (eds.), *Handbook of the Ryukyuan languages*, 113–140. Berlin: de Gruyter. (<https://doi.org/10.1515/9781614511151.113>) (Cited 4.1.5)
- Kaufman, Terry. 1990. Language history in South America: What we know and how to know more. In Payne, Doris L. (ed.), *Amazonian linguistics: Studies in lowland South American languages*, 13–73. Austin: University of Texas Press. (Cited 7.13, Chapter 7, 7.22, 7.29.5)
- Kendon, Adam. 1988. Parallels and divergences between Warlpiri Sign Language and spoken Warlpiri: Analyses of signed and spoken discourses. *Oceania* 58(4). 239–254. (<https://doi.org/10.1002/j.1834-4461.1988.tb02282.x>)
- Kessler, Brett. 2008. The mathematical assessment of long-range linguistic relationships. *Language and Linguistics Compass* 2(5). 821–839. (<https://doi.org/10.1111/j.1749-818X.2008.00083.x>) (Cited 1.1.4)
- King, Alan R. 1994. *The Basque language: A practical introduction*. Reno: University of Nevada Press. (Cited 3.8.5)
- Klima, Edward & Bellugi, Ursula (eds.). 1979. *The signs of language*. Cambridge: Harvard University Press. (Cited 8.9.3)
- Klimov, Georgij Andreevič (Климов, Георгий Андреевич). 1965. *Kavkazskie jazyki* (Кавказские языки) [Caucasian languages]. Moscow: Nauka Publishing House. (Cited 3.3.4)
- Koelle, Sigismund Wilhelm. 1854. *Polyglotta Africana, or a comparative vocabulary of nearly three hundred words and phrases, in more than one hundred distinct African languages*. London: Church Missionary House. (Cited Chapter 5)
- König, Ekkehard & van der Auwera, Johan. 2002. *The Germanic languages*. New York: Routledge. (Cited 3.8.2)
- Kouwenberg, Silvia & Singler, John Victor (eds.). *The handbook of pidgin and creole studies*. Malden: Wiley-Blackwell. (Cited 8.9.1)
- Krauss, Michael E. 1980. *Alaska native languages: Past, present, and future*. Fairbanks: University of Alaska, Alaska Native Language Center. (Cited 7.29.3)
- Krishnamurti, Bhadriraju. 2003. *The Dravidian languages*. New York: Cambridge University Press. (Cited 4.13.8)
- Kroeber, A. L. 1940. Stimulus diffusion. *American Anthropologist* 42(1). 1–20. (<https://doi.org/10.1525/aa.1940.42.1.02a00020>) (Cited 2.2.2)
- Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press. (Cited 6.5.6.1)
- Lefebvre, Claire & Brousseau, Anne-Marie. 2002. *A grammar of Fongbe*. Berlin: de Gruyter. (Cited 5.3.5)

- Levinson, Stephen C. 1997. Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology* 7(1). 98–131. (<https://doi.org/10.1525/jlin.1997.7.1.98>) (Cited 6.3.3)
- Lewis, M. Paul & Simons, Gary F. & Fennig, Charles D. (eds.). 2016. *Ethnologue: Languages of the world*. 19th edn. Dallas: SIL International. (<http://www.ethnologue.com>) (Cited 1.4.1)
- Lhalungpa, Lobsang P. (transl.). 1984. *The life of Milarepa: A new translation from the Tibetan*. Boulder: Shambhala. (Cited 4.11.5)
- Li, Charles N. & Thompson, Sandra A. 1979. Third-person pronouns and zero-anaphora in Chinese discourse. In *Syntax and semantics*, vol. 12, Givón, Talmy (ed.), *Discourse and syntax*, 311–335. New York: Academic. (Cited 6.5.6.3)
- Liddell, Scott K. & Johnson, Robert E. 1989. American Sign Language: The phonological base. *Sign Language Studies* 64. 197–277. (Cited 8.5.1)
- Lipiński, Edward. 2001. *Semitic languages: Outline of a comparative grammar*. 2nd edn. Louvain: Peeters. (Cited 5.9.2)
- Liu, Y. C. 1960. *Fifty Chinese stories, selected from classical texts, romanized and translated into Modern Chinese*. London: Lund Humphries. (Cited 4.10.4)
- Lynch, John. 1998. *Pacific languages: An introduction*. Honolulu: University of Hawai'i Press. (Cited 6.7.1)
- Lynch, John & Ross, Malcolm & Crowley, Terry. 2002. *The Oceanic languages*. Richmond: Curzon. (Cited 6.7.2)
- Mallory, J. P. & Adams, D. Q. 2006. *The Oxford introduction to Proto-Indo-European and the Proto-Indo-European world*. Oxford: Oxford University Press. (Cited 3.8.2)
- Malotki, Ekkehart. 1983. *Hopi time: A linguistic analysis of the temporal concepts in the Hopi language*. Berlin: Walter de Gruyter. (Cited 7.12)
- Martin, Samuel E. 1962. *Essential Japanese: An introduction to the standard colloquial language*. 3rd rev. edn. Rutland: Tuttle. (Cited 1.2.2)
- Matisoff, James A. 1991. Sino-Tibetan linguistics: Present state and future prospects. *Annual Review of Anthropology* 20. 469–504. (<https://doi.org/10.1146/annurev.an.20.100191.002345>) (Cited 4.13.4)
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press. (Cited 4.3.2, 4.13.4)
- Matras, Yaron, & Bakker, Peter (eds.). 2003. *The mixed language debate: Theoretical and empirical advances*. Berlin: de Gruyter. (Cited 8.9.1)
- Max Planck Institute for Evolutionary Anthropology. 2008. Leipzig glossing rules. (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>) (Cited 2.1, 1.2.1)
- McBurney, Susan. 2012. History of sign languages and sign language linguistics. In Pfau, Roland & Steinbach, Markus & Woll, Bencie (eds.), *Sign language: An international handbook*, 909–948. Berlin: de Gruyter. (<https://doi.org/10.1515/9783110261325.909>) (Cited 8.5.1)
- McMahon, April & McMahon, Robert. 2005. *Language classification by numbers*. New York: Oxford University Press. (Cited 1.4.5)
- McManus, Damian. 1991. *A guide to ogam*. Maynooth: An Sagart. (Cited 3.1.1)
- Meinhof, Carl. 1906. Grundzüge einer vergleichenden Grammatik der Bantusprachen [Outline of a comparative grammar of the Bantu languages]. Berlin: Dietrich Reimer. (Cited 5.7.3.1)

- Meir, Irit & Sandler, Wendy & Padden, Carol & Aronoff, Mark. 2010. Emerging sign languages. In Marschark, Marc & Spencer, Patricia Elizabeth (eds.), *The Oxford handbook of deaf studies, language, and education*, vol. 2, 267–280. Oxford: Oxford University Press. (<https://doi.org/10.1093/oxfordhb/9780195390032.013.0018>) (Cited 8.5.2)
- Melchert, H. Craig. In press. The position of Anatolian. In Weiss, Michael & Garrett, Andrew (eds.), *Handbook of Indo-European Studies*, Chapter 6. Oxford University Press. (Cited 3.1.9)
- Merrifield, William B. & Naish, Constance M. & Rensch, Calvin R. & Story, Gillian (eds.). 1987. *Laboratory manual for morphology and syntax*. 4th rev. edn. Santa Ana: Summer Institute of Linguistics. (Cited Tab. 6.13, 7.28.2–3)
- Mihalic, F. 1971. *The Jacaranda dictionary and grammar of Melanesian Pidgin*. Brisbane: Jacaranda Press. (Cited 8.7.2.1, 8.8.1)
- Millar, Robert McColl. 2015. *Trask's historical linguistics*. 3rd edn. London: Routledge. (Cited 1.4.4)
- Mistry, P. J. 1996. Gujarati writing. In Daniels, Peter T. & Bright, William (eds.), *The world's writing systems*, 391–394. New York: Oxford University Press. (Cited 2.1.3.4)
- MIT-Haiti. 2015. Pi bon zouti pou nou rebati [Best practices for reconstruction]. (<http://haiti.mit.edu/ht>) (Accessed 2016-05-28.) (Cited 8.4)
- Mithun, Marianne. 1996. Overview of general characteristics. In Goddard, Ives (ed.), *Handbook of North American Indians*, vol. 17, *Languages*, 137–157. Washington: Smithsonian Institution. (Cited 7.29.2)
- Mithun, Marianne. 1999. *The languages of native North America*. Cambridge: Cambridge University Press. (Cited Chapter 7, 7.29.2)
- Miyaoka, Osahito. 2012. *A grammar of Central Alaskan Yupik (CAY)*. Berlin: De Gruyter Mouton. (Cited 7.26.3.4)
- Mohamed, M. A. 2001. *Modern Swahili grammar*. Nairobi: East African Educational Publishers. (Cited 5.9.7)
- Montgomery-Anderson, Brad. 2008. *A reference grammar of Oklahoma Cherokee*. Lawrence: University of Kansas. (Doctoral dissertation). (http://kuscholarworks.ku.edu/dspace/bitstream/1808/4212/1/umi-ku-2613_1.pdf) (Cited 2.1.3.2, 7.6)
- Mora-Marín, David F. 2008. Full phonetic complementation, semantic classifiers, and semantic determinatives in Ancient Mayan hieroglyphic writing. *Ancient Mesoamerica* 19. 195–213. (<https://doi.org/10.1017/S0956536108000345>) (Cited 2.2.2)
- Moseley, Christopher (ed.). 2010. *Atlas of the world's languages in danger*. 3rd edn. Paris: UNESCO Publishing. (<http://www.unesco.org/culture/en/endangeredlanguages/atlas>) (Cited 3.2.2, 6.5.1, 8.6.3)
- Mous, Maarten. 2003. *The making of a mixed language: The case of Ma'a/Mbugu*. Amsterdam: Benjamins. (<https://doi.org/10.1075/cll.26>) (Cited 8.9.1)
- Mufwene, S. 2001. Pidgins and creoles. In Smelser, Neil J. & Baltes, Paul B. (eds.), *International encyclopedia of the social & behavioral sciences*, 11440–11445. Kidlington: Elsevier. (<https://doi.org/10.1016/Bo-08-043076-7/02939-9>) (Cited 8.9.1)
- MultiTree: A digital library of language relationships. 2014. Bloomington: Indiana University Department of Linguistics. (<http://multitree.org>) (Cited 1.4.1)
- Mutaka, Ngessimo N. 2000. *An introduction to African linguistics*. (With the collaboration of Pius Ngwa Tamanji.) Munich: Lincom. (Cited 5.9.4)

- Muysken, Pieter. 1997. Media Lengua. In Thomason, Sarah G. (ed.), *Contact languages: A wider perspective*, 365–426. Amsterdam: Benjamins. (<https://doi.org/10.1075/cll.17>) (Cited 8.1)
- Myachina, E. N. 1981. *The Swahili language: A descriptive grammar*. Boston: Routledge and Kegan Paul. (Translated from the original Russian by G. L. Campbell.) (Cited 5.7.3.1)
- Nguyễn, Đình-Hoà. 1997. *Vietnamese*. Philadelphia: Benjamins. (<https://doi.org/10.1075/loall.9>) (Cited 4.13.7)
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press. (Cited 1.4.6)
- Nichols, Johanna. 2007. What, if anything, is typology? *Linguistic Typology* 11(1). 231–238. (<https://doi.org/10.1515/LINGTY.2007.017>) (Cited 1.4.6)
- Nichols, Johanna. 2013. Causativization and contact in Nakh-Daghestanian. In Cathcart, Chundra & Kang, Shinae & Sandy, Clare S. (eds.), *Proceedings of the 37th annual meeting of the Berkeley Linguistics Society: Special session on languages of the Caucasus*, 68–80. Berkeley: Berkeley Linguistics Society. (Cited 3.3.2)
- Nichols, Johanna & Warnow, Tandy. 2008. Tutorial on computational linguistic phylogeny. *Language and Linguistics Compass* 2. 760–820. (<https://doi.org/10.1111/j.1749-818X.2008.00082.x>) (Cited 1.4.5)
- Norman, Jerry. 1988. *Chinese*. New York: Cambridge University Press. (Cited 4.13.4)
- Nugteren, Hans. 2011. *Mongolic phonology and the Qinghai–Gansu languages*. Leiden: Universiteit Leiden. (Doctoral dissertation.) (Cited 4.1.2)
- O'Connor, M. 1996. Epigraphic Semitic scripts. In Daniels, Peter T. & Bright, William (eds.), *The world's writing systems*, 88–107. New York: Oxford University Press. (Cited 2.1.3.3)
- Odden, David. 1982. Tonal phenomena in Kishambaa. *Studies in African Linguistics* 13(2). 177–208. (Cited 5.3.7)
- Osborne, C. R. 1974. *The Tiwi language*. Canberra: Australian Institute of Aboriginal Studies. (Cited 6.3.2)
- Owino, Daniel. 2003. *Phonological nativization of Dholuo loanwords*. Pretoria: University of Pretoria. (Doctoral thesis.) (Cited 5.2.1)
- Parker, Gary J. 1969. *Ayacucho Quechua grammar and dictionary*. The Hague: Mouton. (Cited 7.27.3.4, 7.27.5)
- Patrie, James. 1982. *The genetic relationship of the Ainu language*. Honolulu: University of Hawai'i Press. (Cited 4.13.2)
- Pawley, Andrew & Attenborough, Robert & Golson, Jack & Hide, Robin (eds.). 2005. *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics, Australian National University. (Cited 6.7.3)
- Payne, Doris Lander. 1985. *Aspects of the grammar of Yagua: A typological perspective*. Los Angeles: University of California. (Doctoral dissertation.) (Cited 7.20)
- Payne, Thomas E. 1980. Who's kissing who: The fourth person in Yup'ik Eskimo. In Munro, Pamela (ed.), *Studies of switch-reference*, 65–88. Los Angeles: University of California. (Cited 7.26.3.4)
- Peirce, Charles S. 1984. *Writings of Charles S. Peirce: A chronological edition*, vol. 2. Bloomington: Indiana University Press. (Cited 2.1)
- Perrott, D. V. 1951. *Teach yourself Swahili*. London: English Universities Press. (Cited 5.7.5, 5.8.1)
- Pfau, Roland & Steinbach, Markus & Woll, Bencie (eds.). 2012. *Sign language: An international handbook*. Berlin: de Gruyter. (<https://doi.org/10.1515/9783110261325>) (Cited 8.9.3)

- Polivanov, E. D. 1916. Indo-evropejskoe *medhu—obshche-kitajskoe *mit (Индоевропейское *medhu—общекитайское *mit) [Indo-European *medhu is Common Chinese *mit]. *Zapiski Vostočnogo Otdelenija Rossijskogo Arхеологического Общества* (Записки Восточного отделения Российского археологического общества) 2. 263–264. (Cited 3.1.10)
- Polomé, Edgar C. 1967. *Swahili language handbook*. Washington: Center for Applied Linguistics. (Cited 5.8.4)
- Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press. (Cited 4.10.4)
- Ramos, Teresita V. 1971. *Tagalog structures*. Honolulu: University of Hawai'i Press. (Cited 6.6.2)
- Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics. (Cited 4.4, 4.13.5)
- Reed, Irene. 1977. *Yup'ik Eskimo grammar*. Fairbanks: University of Alaska, Alaska Native Language Center. (Cited 7.26.1)
- Roberts, Dexter. 2014. 400 million Chinese can't speak Mandarin and Beijing is worried. (<http://www.bloomberg.com/bw/articles/2014-09-23/say-that-in-mandarin-please>) (Cited 4.10.1)
- Robinson, Andrew. 2007. *The story of writing: Alphabets, hieroglyphs, and pictograms*. New edn. New York: Thames and Hudson. (Cited 2.4)
- Rogers, Henry. 2004. *Writing systems: A linguistic approach*. Malden: Blackwell. (Cited 2.4)
- Róna-Tas, András. 2007. Nutshell Chuvash. (http://www.academia.edu/10108012/Erasmus_Mundus_Intensive_Program_Turkic_languages_and_cultures_in_Europe_TLCE_Nutshell_Chuvash) (Cited 4.1.1)
- Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Pawley, Andrew & Attenborough, Robert & Golson, Jack & Hide, Robin (eds.), *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–65. Canberra: Pacific Linguistics. (Cited 6.2)
- Rubin, Aaron D. 2008. The subgrouping of the Semitic languages. *Language and Linguistics Compass* 2(1). 61–84. (<https://doi.org/10.1111/j.1749-818x.2007.00044.x>) (Cited 5.1.1)
- Rude, Noel. 1986. Topicality, transitivity, and the direct object in Nez Perce. *International Journal of American Linguistics* 52(2). 124–153. (Cited 7.9)
- Salminen, Tapani. 2002. Problems in the taxonomy of the Uralic languages in the light of modern comparative studies. In *Lingvističeskij bespredel: Sbornik statej k 70-letiju A. I. Kuznetsovoj* (Linguistic lawlessness: Collection of articles on A. I. Kuznetsova's 70th birthday), 44–55. Moscow: Izdatel'stvo Moskovskogo Universiteta. (Cited 3.2.1)
- Salmons, Joseph C. & Joseph, Brian D. (eds.). 1997. *Nostratic: Sifting the evidence*. Amsterdam: Benjamins. (<https://doi.org/10.1075/cilt.142>) (Cited 4.1.7)
- Saltarelli, Mario. 1988. *Basque*. London: Croom Helm. (Cited 1.2.2)
- Sampson, Geoffrey. 2015. *Writing systems*. 2nd edn. Sheffield: Equinox. (Cited 2.4)
- Sandler, Wendy & Lillo-Martin, Diane. 2006. *Sign language and linguistic universals*. Cambridge: Cambridge University Press. (Cited 8.9.3)
- Sandler, Wendy & Meir, Irit & Padden, Carol & Aronoff, Mark. 2005. The emergence of grammar: Systematic structure in a new language. *PNAS* 102(7). 2661–2665. (<https://doi.org/10.1073/pnas.0405448102>) (Cited 8.5.2)

- Sapir, Edward. 1915. The Na-Dene languages: A preliminary report. *American Anthropologist* 17(3). 534–558. (<https://doi.org/10.1525/aa.1915.17.3.02a00080>) (Cited 7.2)
- Schieffelin, Bambi B. & Charlier Doucet, Rachele. 1992. The “real” Haitian Creole: Metalinguistics and orthographic choice. *Pragmatics* 2(3). 427–443. (Cited 8.4)
- Schmandt-Besserat, Denise. 2002. Signs of life. *Archaeology Odyssey* 5(1). 1–3. (Cited 2.2.1.4, 2.4)
- Schmidt, Annette. 1985. *Young people’s Dyirbal: An example of language death from Australia*. Cambridge: Cambridge University Press. (Cited 6.5.1)
- Schütz, Albert J. 1980. A reanalysis of the Hawaiian vowel system. *Oceanic Linguistics* 20(1). 1–43. (<https://doi.org/10.2307/3622840>) (Cited 6.4.2)
- Schütz, Albert J. 1994. *The voices of Eden: A history of Hawaiian language studies*. Honolulu: University of Hawai‘i Press. (Cited 6.7.6)
- Sears, Richard Horace. 2016. Chinese etymology. (<http://www.chineseetymology.org>) (Cited 2.1.2)
- Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18(3). 445–479. (Cited 5.3.5)
- Shibatani, Masayoshi. 1990. *The languages of Japan*. Cambridge: Cambridge University Press. (Cited 4.1.6, 4.13.2)
- Siegel, Jeff. 2008. *The emergence of pidgin and creole languages*. Oxford: Oxford University Press. (Cited 8.9.1)
- SIL International. 2015. ISO 639-3 Registration Authority. (<http://www-01.sil.org/iso639-3>) (Cited 2.4, Chapter 3)
- Singh, Ishtla. 2000. *Pidgins and creoles: An introduction*. New York: Routledge. (Cited 8.9.1)
- Snyman, J. W. 1970. *An introduction to the !Xū (!Kung) language*. Cape Town: A. A. Balkeme. (Cited 5.8.2)
- Sohn, Ho-Min. 2001. *The Korean language*. New York: Cambridge University Press. (Cited 4.13.2)
- Solzhenitsyn, Aleksandr. 1968. *V pervom krugu* (В первом кругу) [In the first circle]. London: Flegon. (Cited 3.5.6)
- Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax*. Harlow: Pearson Education. (Cited 1.4.6)
- Song, Jae Jung (ed.). 2011. *The Oxford handbook of linguistic typology*. Oxford: Oxford University Press. (Cited 1.4.6)
- Starostin, George. 2010. Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3. 79–116. (Cited 1.4.5)
- Starostin, George. 2012. Dene-Yeniseian: A critical assessment. *Journal of Language Relationship* 8. 117–152. (Cited 4.2)
- Starostin, Sergei & Dybo, Anna & Mudrak, Oleg. 2003. *Etymological dictionary of the Altaic languages*. Leiden: Brill. (Cited 4.1.7)
- Stokoe, William C., Jr. 1960. *Sign language structure: An outline of the visual communication systems of the American deaf*. (Studies in Linguistics, Occasional Papers, 8.) Buffalo: Dept. of Anthropology and Linguistics, University of Buffalo. (Cited 8.5.1, 8.9.3)
- Sun, Chaofen. 2006. *Chinese: A linguistic introduction*. New York: Cambridge University Press. (Cited 4.13.4)
- Sussex, Roland & Cubberley, Paul. 2011. *The Slavic languages*. Cambridge: Cambridge University Press. (Cited 3.8.2)

- Swadesh, Morris. 1955. Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21(2). 121–137. (Cited 1.1.4)
- Tadmor, Uri & Haspelmath, Martin & Taylor, Bradley. 2012. Borrowability and the notion of basic vocabulary. In Wichmann, Søren & Grant, Anthony P. (eds.), *Quantitative approaches to linguistic diversity: Commemorating the centenary of the birth of Morris Swadesh*, 35–55. Amsterdam: Benjamins. (<https://doi.org/10.1075/bct.46>) (Cited 8.1)
- Thomason, Sarah Grey (ed.). 1997. *Contact languages: A wider perspective*. New York: Benjamins. (<https://doi.org/10.1075/cll.17>) (Cited 8.9.1)
- Thomason, Sarah Grey & Kaufman, Terrence. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press. (Cited 8.9.1)
- Thurgood, Graham. 1993. Tai-Kadai and Austronesian: The nature of the historical relationship. *Oceanic Linguistics* 33(2). 345–368. (<https://doi.org/10.2307/3623133>) (Cited 4.5)
- Thurgood, Graham & LaPolla, Randy J. (eds.). 2003. *The Sino-Tibetan languages*. New York: Routledge. (Cited 4.13.4)
- Tolstoy, Leo. 1852. Detstvo (Детство) [Childhood]. (<http://www.klassika.ru/read.html?proza/tolstoj/detstwo.txt>) (Cited 3.7.1)
- Tournadre, Nicolas. 2005. Laire linguistique tibétaine et ses divers dialectes [The Tibetan linguistic area and its various dialects]. *Lalies* 25. 7–56. (Cited 4.11.1)
- Tranter, Nicolas (ed.). 2012. *The languages of Japan and Korea*. New York: Routledge. (Cited 4.13.1)
- Treiman, Rebecca & Kessler, Brett. 2014. *How children learn to write words*. Oxford: Oxford University Press. (Cited 4.1.4)
- Tritton, A. S. 1943. *Teach yourself Arabic*. Philadelphia: D. McKay. (Cited 5.6.6)
- Unger, J. Marshall. 1990. Japanese and what other Altaic languages? In Baldi, Philip (ed.), *Linguistic change and reconstruction methodology*, 547–562. New York: de Gruyter. (Cited 4.13.2)
- Vajda, Edward J. 2011. A Siberian link with Na-Dene languages. In Kari, James & Potter, Ben A. (eds.), *The Dene-Yeniseian connection*, 33–99. Fairbanks: University of Alaska Department of Anthropology and Alaska Native Language Center. (Cited 4.2)
- van Eijk, Jan P. 1998. The morphological status of partial reduplication: Evidence from Lushootseed and Lillooetvan. In Janse, Mark (ed.), *Productivity and creativity: Studies in general and descriptive linguistics in honor of E. M. Uhlenbeck*, 433–448. Berlin: de Gruyter. (<https://doi.org/10.1515/9783110806946.433>) (Cited 7.13)
- Venezky, Richard L. 1999. *The American way of spelling: The structure and origins of American English orthography*. New York: Guilford Press. (Cited 4.11.2)
- Verhaar, John W. M. 1995. *Towards a reference grammar of Tok Pisin: An experiment in corpus linguistics*. Honolulu: University of Hawai'i Press. (Cited 8.9.2)
- Vovin, Alexander. 2005. The end of the Altaic controversy: In memory of Gerhard Doerfer. *Central Asiatic Journal* 49(1). 71–132. (Review of Starostin et al. 2003.) (Cited 4.1.7)
- Walker, Rachel Leah. 1998. *Nasalization, neutral segments, and opacity effects*. Santa Cruz: University of California. (Doctoral dissertation.) (Cited 7.24)
- Wáng, Fùshì (王辅世) (ed.). 1985. *Miáoyǔ jiǎnzhi* (苗语简志) [A sketch of Miao]. Beijing: Mǐnzú Chūbǎnshè. (Cited 4.12.4)
- Watkins, Laurel J. 1976. Position in grammar: Sit, stand, lie. *Kansas Working Papers in Linguistics* 1. 16–41. (<https://doi.org/10.17161/KWPL.1808.702>) (Cited 7.4)

- Webb, Vic, & Kembo-Sure. 2000. *African voices: An introduction to the languages and linguistics of Africa*. Oxford: Oxford University Press. (Cited 5.9.1)
- Whaley, Lindsay J. 1997. *Introduction to typology: The unity and diversity of language*. Newbury Park: Sage. (Cited 1.4.6)
- Whitney, Arthur H. 1956. *Finnish*. Sevenoaks: Hodder and Stoughton. (Cited 3.8.7)
- Whorf, Benjamin Lee. 1956. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. (Carroll, John B., ed.) Cambridge: MIT Press. (Cited 7.12)
- Wichmann, Søren. 1995. *The relationship among the Mixe-Zoquean languages of Mexico*. Salt Lake City: University of Utah Press. (Cited 7.16)
- Williamson, Kay & Blench, Roger. 2000. Niger-Congo. In Heine, Bernd & Nurse, Derek (eds.), *African languages: An introduction*, 11–42. Cambridge: Cambridge University Press. (Cited 5.3)
- Wonderly, William L. & Gibson, Lorna F. & Kirk, Paul L. 1954. Number in Kiowa: Nouns, demonstratives, and adjectives. *International Journal of American Linguistics* 20(1). 1–7. (Cited 7.11)
- Wurm, Stephen A. 1982. *Papuan languages of Oceania*. Tübingen: Narr. (Cited 6.7.3)
- Xiao, Wen & Treiman, Rebecca. 2012. Iconicity of simple Chinese characters. *Behavior Research Methods* 44(4). 954–960. (<https://doi.org/10.3758/s13428-012-0191-3>) (Cited 2.1.2)
- Yamamoto, Akira Y. & Zepeda, Ofelia. 2004. Native American languages. In Finegan, Edward & Rickford, John R. (eds.), *Language in the USA: Themes for the twenty-first century*, 153–181. New York: Cambridge University Press. (Cited 7.29.2)
- Young, Mark C. (ed.) 1996. *The Guinness Book Of World Records, 1997*. Enfield: Guinness World Publishing. (Cited 3.3.1–2)
- Zariquiey, Roberto & Córdova, Gavina. 2008. Qayna, kunan, paqarin: Una introducción práctica al quechua chanca [Yesterday, today, tomorrow: A practical introduction to Chanca Quechua]. San Miguel: Pontificia Universidad Católica del Perú. (Cited 7.27.3.5)
- Zhōngyāng Mǐnzú Xuéyuàn. Shǎoshǔ Mǐnzú Yǔyán Yánjiùsuǒ (中央民族学院少数民族语言研究所). 1987. *Zhōngguó shǎoshǔ mǐnzú yǔyán* (中国少数民族语言) [The languages of China's minorities]. Chengdu: Sichuān Mǐnzú Chūbǎnshè. (Cited 4.12.1)
- Zúñiga, Fernando. 2006. Mapudungun, el habla mapuche: Introducción a la lengua mapuche, con notas comparativas y un CD [Mapudungun, the Mapuche speech: Introduction to the Mapuche language, with comparative notes and a CD]. Santiago de Chile: Centro de Estudios Públicos. (Cited 7.23)

LANGUAGE INDEX

- Abaza (ABQ), 75–76
 Abkhaz (ABK), 75–76
 Abkhaz-Abaza, 75
 Achuar-Shiwiari (ACU), 352, 356
 Achumawi (ACV), 326, 328, 338
 Adamawa-Ubangi, 214
 Adamorobe Sign Language (ADS), 416, 420
 Adi (ADI), 138, 141
 Adyghe (ADY), 75–76
 Afar (AAR), 199, 205
 Afrikaans (AFR), 60, 213, 222
 Afro-Asiatic, 136, 149, 197–208, 218, 223, 244
 Aguaruna (AGR), 352, 356
 A-Hmao (HRM), 138, 143
 Aikanā (TBA), 352, 360
 Ainu (AIN), 128–130, 134–137
 Akan (AKA), 213–215
 Akateko (KNJ), 345, 350
 Akawaio (AKE), 352, 363
 Akha (AHK), 138, 141
 Akkadian (AKK), 198–202
 Alabama (AKZ), 328, 335
 Alabama-Koasati, 335
 Albanian (SQI), 56–57, 65
 Albanian Sign Language (SQK), 416
 Aleut (ALE), 327–328, 403
 Algerian Arabic (ARQ), 224
 Algerian Sign Language (ASP), 416
 Algic, 326, 332–336
 Algonquian, 327, 332–335, 344
 Algonquin (ALQ), 333
 Al-Sayyid Bedouin Sign Language (SY),
 416, 419
 Altai (ALT), 128, 130
 Altaic, 127–136, 194, 270
 Alutor (ALR), 128, 136
 Alywarr (ALY), 266, 275–276
 Amahuaca (AMC), 352, 357
 Amara (AMR), 352, 355
 Amdo (ADX), 138, 167–168
 American Sign Language (ASE), 416–418, 420,
 423, 448–449
 Amerind, 344, 351
 Amharic (AMH), 49, 198–201
 Amis (AMI), 266–267
 Amuzgo (AMU), 346
 Amuzgo-Mixtecan, 346
 Anatolian, 56, 69
 Ancient Greek (GRC), 9, 60, 64, 69
 Andamanese, 265, 275
 Andi (ANI), 75–77
 Andoque (ANO), 352, 356
 Angami (NJM), 138, 141
 Anglo-Frisian, 60
 Anindilyakwa (AOI), 266, 275–276
 Anmatyerre (AMX), 266, 275–276
 Ao (NJO), 138, 141
 Apurinā (APU), 352, 355
 Apurucayali (CPC), 352, 355
 Arabela (ARL), 352, 356
 Arabic (ARA), 10, 15, 31, 45, 68, 79, 149,
 198–203, 205–206, 211, 222–242, 244, 246,
 274, 411, 413–414, 419
 Arabic dialects, 223
 Aramaic (ARC), 149, 200–201, 225
 Arandic, 275–276
 Arapaho (ARP), 328, 333
 Arauan, 355
 Arawak (ARW), 352, 354
 Arawakan, 352, 354, 362, 402
 Argentine Sign Language (AED), 416
 Arikapú (ARK), 352, 360
 Arikara (ARI), 328, 338
 Arikem, 361
 Armenian (HYE), 56–57, 67, 70
 Armenian Sign Language (AEN), 416
 Arutani (ATX), 352, 363
 Asante (TWI), 215
 Asha-Ashe, 355
 Asha-Ashe-Kak-Matsi-Nan, 355
 Asháninka (CNI), 352, 355
 Ashanti, 215
 Ashéninka, 355
 Ashéninka Pajonal (CJO), 352, 355
 Aslian, 146
 Asoa (ASV), 199, 209, 211
 Assamese (ASM), 413
 Assiniboine (ASB), 328, 336
 Atayal (TAY), 266–267
 Atayalic, 267
 Athabaskan, 330–332, 331
 Athabaskan-Eyak-Tlingit, 330

- Atlantic, 214–215, 218
 Atsi (ATB), 138, 141
 Aukan (DJK), 413
 Australian, 265, 274–278, 306
 Australian Sign Language (ASF), 416
 Austrian Sign Language (ASQ), 416
 Austric, 270
 Austroasiatic, 127, 145–147, 195, 270
 Austronesian, 11, 13, 145, 149, 197, 222,
 265–272, 266, 272, 274, 276, 278, 319–320,
 330, 428–429
 Austro-Tai, 145, 270
 Avar (AVA), 75–77
 Avar-Andic, 77
 Avestan (AVE), 57, 67, 69–70, 103
 Awa-Cuaiquer (KWI), 352–353
 Awakateko (AGU), 345, 350
 Awakateko-Ixil, 350
 Awetí (AWE), 352, 361
 Awjilah (AUJ), 199, 203
 Ayacucho Quechua (QUY), 352, 356, 372–394
 Aymara (AYR), 352, 356, 373
 Aymaran, 356
 Ayoreo (AYO), 352, 359
 Azerbaijani (AZB), 81, 128, 130
 Aztecán, 341–342
 Aztec-Tanoan, 343
- Babine (BCR), 328, 331
 Babine-Carrier, 331
 Bahamas Creole English (BAH), 413
 Bahasa Indonesia (IND), 269, 274
 Bahasa Malaysia, 269
 Bahnar (BDQ), 138, 146
 Bahnaric, 146
 Bai (BCA), 138, 140–141
 Balinese (BAN), 266–267, 269
 Balochi (BAL), 57, 67, 148
 Balti (BFT), 138, 167
 Baltic, 65, 74
 Balto-Slavic, 25, 56, 65–67, 70–71
 Bamako Sign Language (BOG), 416, 420
 Bambara (BAM), 213–214
 Bantoid, 216, 218
 Bantoid-Cross, 216
 Bantu, 216–218, 220–222, 241, 243–244, 403
 Baoulé (BCI), 213–214
 Barbacoan, 351, 353
 Barish, 141
 Bashkir (BAK), 128, 130
 Basque (EUS), 7–8, 24, 55–56, 79–80, 148
 Batsbi (BBL), 75–77
 Belarusian (BEL), 57, 65, 81
 Belize Kriol (BZI), 413
 Bella Coola (BLC), 328, 343–344
- Bengali (BEN), 57, 67–68
 Bengkulu Sign Language (BQY), 416
 Benue-Congo, 214, 216
 Berber, 198, 200, 202–204, 208
 Berta (WTI), 199, 209
 Bilua (BLB), 266, 273
 Bislama (BIS), 413
 Biu-Mandara, 204
 Blackfoot (BLA), 328, 333–335, 357
 Bodo (BRX), 138, 141
 Bolivian Sign Language (BVL), 416
 Bolivia-Parana Arawakan, 355
 Bora (BOA), 352, 356
 Boran, 356
 Bororo (BOR), 352, 359
 Bororoan, 359
 Brahui (BRH), 138, 147–148
 Brazilian Sign Language (BZS), 416
 Breton (BRE), 57–58, 427
 Bribri (BZD), 352–353
 British Sign Language (BFI), 416–417, 423
 Brythonic, 58
 Buang (BZH), 320
 Buginese (BUG), 269
 Bulgarian (BUL), 57, 65–66, 81
 Bunuban, 275–276
 Bura (BWR), 199, 204
 Bura Sign Language, 420
 Buriat (BUA), 128, 130
 Burmese (MYA), 138, 140–142
 Burmish, 141
 Burushaski (BSK), 80, 127–128, 148–149, 245
- Caddo (CAD), 328, 338
 Caddoan, 326, 338
 Cahitan, 342
 Cahuapanan, 353, 356
 Cahuilla (CHL), 328, 342
 Cameroonian Pidgin English (WES), 409
 Camsá (KBH), 352–353
 Canaanite, 200–201
 Candoshi-Shapra (CBU), 352, 356
 Cantonese, 140, 405–406
 Capanahua (KAQ), 352, 357
 Car (CAQ), 138, 146
 Carib (CAR), 352, 363
 Cariban, 362–363
 Caribbean Arawakan, 354
 Carrier (CRX), 328, 331
 Cashibo-Cacataibo (CBR), 352, 357
 Catalan (CAT), 57, 64
 Caucasus, 75–79, 148
 Cavineña (CAV), 352, 357
 Cayuga (CAY), 328, 337
 Cebuano (CEB), 266–268

- Celebic, 269
 Celtic, 56, 58–59, 70–71, 426–427
 Central Alaskan Yup'ik (ESU), 17, 21–22, 328,
 330, 363–372
 Central Cariban, 363
 Central Dravidian, 147
 Central-Eastern Austronesian, 267
 Central-Eastern Malayo-Polynesian, 269–270
 Central Jean, 359
 Central Kanuri (KNC), 211
 Central Malayo-Polynesian, 269
 Central Mande, 214
 Central M̄in (CZO), 138–139
 Central Mongolic, 130
 Central Numic, 342
 Central Pacific, 270
 Central Pame (PBS), 345, 347
 Central Quechua, 373
 Central Salishan, 343
 Central Semitic, 200–201, 223
 Central Siberian Yupik (ESS), 328, 330, 364
 Central Sudanic, 209–211
 Central Tibetan (BOD), 138, 167–168, 188
 Central West Gurage (SGW), 199–200
 Central Zoque (ZOC), 345, 348
 Chachi (CBI), 352–353
 Chácobo (CAO), 352, 357
 Chadian Sign Language (CDS), 416
 Chadic, 200, 204–205
 Cham (CJA), 266–267, 269
 Chamacoco (CEG), 352, 359
 Chamic, 149
 Chamorro (CHA), 266–268, 270
 Chanka, 373
 Chapacuran, 354
 Chatino, 347
 Chavacano (CBK), 413
 Chayahuita (CBT), 352, 356
 Chechen (CHE), 75–77
 Cherokee (CHR), 328, 337
 Cheyenne (CHY), 328, 333
 Chiangmai Sign Language (CSD), 416
 Chiapas Zoquean, 348
 Chibchan (CBA), 344, 351, 353
 Chichimeca-Jonaz (PEI), 345, 347
 Chickasaw (CIC), 328, 335
 Chilcotin (CLC), 328, 331
 Chilean Sign Language (CSG), 416
 Chimalapa Zoquean, 348
 Chimborazo Highland Quichua (QUG), 352,
 356, 373
 Chinantecan, 347
 Chinese (ZHO), 32, 35–38, 40, 131–132, 137,
 139–140, 142, 144–146, 149, 168, 217, 268,
 274, 406, 415, 424
 Chinese Pidgin English (CPI), 405–409, 412
 Chinese Sign Language (CSL), 416
 Chinookan, 326, 408–409
 Chinook Jargon (CHN), 407–409
 Chipaya (CAP), 352, 356
 Chipaya-Uru, 356
 Chipewyan (CHP), 328, 331
 Chippewa (CIW), 333
 Chiquitano (CAX), 352, 359
 Choapan (ZPC), 345, 347
 Chocoan, 351, 353
 Choctaw (CHO), 328, 335, 409
 Chol (CTU), 345, 350
 Cholan, 349–350
 Cholan-Tzeltalan, 350
 Chol-Chontal, 350
 Chontal (CHF), 345, 350
 Chuanqiandian (CQD), 143
 Chuj (CAC), 345, 350
 Chujean, 350
 Chukchi (CKT), 128, 136–137
 Chukotko-Kamchatkan, 136
 Chuvash (CHV), 128, 130–131
 Circassian, 75
 Classical Arabic (ARB), 223
 Classical Nahuatl (NCI), 328, 341–342
 Classical Tibetan (XCT), 140, 165–188, 302
 Coast Tsimshian, 328
 Cocama-Cocamilla (COD), 352, 360
 Cochimí (COI), 338
 Cochimí-Yuman, 338
 Cocopa (COC), 328, 338
 Coeur d'Alene (CRD), 328, 343
 Cofán (CON), 352, 356
 Colombian Sign Language (CSN), 416
 Columbian (COL), 328, 343
 Comanche (COM), 328, 342
 Comox (COO), 328, 343
 Coptic (COP), 199, 205–206
 Cora (CRN), 328, 342
 Corachol, 342
 Core K'ichean, 350
 Core Mayan, 350
 Cornish (COR), 57–58, 426
 Costanoan, 326
 Costa Rican Sign Language (CSR), 416
 Coyutla Totonac (TOC), 345–346, 348
 Cree (CRE), 328, 333, 402
 Creek, 335
 Crimean Tatar (CRH), 128, 130
 Crioulo (POV), 413
 Croatian (HRV), 66
 Croatian Sign Language (CSQ), 416
 Cross River, 216–217
 Crow (CRO), 328, 336

- Cuban Sign Language (CSF), 416
 Cubeo (CUB), 352–353, 355, 362
 Cuicatec (CUX), 345–346
 Curripaco (KPC), 352, 355
 Cusco Quechua (QUZ), 352, 356, 373–374
 Cushitic, 200, 205–206, 221, 403
 Czech (CES), 25, 57, 65
 Czech Sign Language (CSE), 416
- Daghestanian, 75, 77
 Dagur (DTA), 128, 130
 Dahalo (DAL), 221
 Dakota (DAK), 328, 336
 Dakotan, 336
 Damin, 404
 Danish (DAN), 57, 60, 62, 327, 346, 365
 Dargwa (DAR), 75–77
 Dari (PRS), 68, 229
 Dbus, 168
 Defaka (AFN), 213–214
 Delta-Californian, 338
 Dene, 331
 Dëne Sųliné, 331
 Dhivehi (DIV), 68
 Dholuo (LUO), 199, 209–210
 Dhuwal (DUJ), 266, 275–276
 Dimasa (DIS), 138, 141
 Dinka (DIN), 199, 209–210
 Dogon (DTS), 213–214
 Dogon-Congo, 214
 Dogrib (DGR), 328, 331
 Dong (KMC), 138, 144
 Dravidian, 127, 145, 147–148, 195, 277
 Duruwa (PCI), 138, 147
 Dutch (NLD), 4, 57, 60, 80, 150, 222, 278
 Dyirbal (DBL), 266, 275, 277, 298–318, 334
 Dzongkha (DZO), 138, 167
 Dzongkha-Lhokä, 167–168
- East Benue-Congo, 215–218, 241
 East Caucasian, 75–77, 148
 East Chadic, 204
 Eastern Algonquian, 333
 Eastern Arabic, 223
 Eastern Arrernte (AER), 266, 275–276
 Eastern Austroasiatic, 146
 Eastern Berber, 203
 Eastern Bolivian Guaraní (GUI), 352, 360
 Eastern Highland (OTM), 345, 347
 Eastern Huasteca Nahuatl (NHE), 328, 342
 Eastern M̐n (CDO), 138–139
 Eastern Oto-Manguean, 346
 Eastern Polynesian, 267, 270
 Eastern Pwo (KJP), 138, 141
 Eastern Sinitic, 139
- Eastern Slavic, 65
 Eastern Sudanic, 209–210
 East Germanic, 62
 East Gurage (STV), 199–200
 East Scandinavian, 60, 62
 East Semitic, 200
 East Slavic, 67, 80, 104
 Ecuadorian Sign Language (ECS), 416
 Edo (BIN), 213, 216–217
 Efe (EFE), 199, 209, 211
 Efik (EFI), 213, 216–218
 Egyptian (EGY), 40, 48, 199–200, 205–208
 Egyptian Arabic (ARZ), 223–224
 Emberá-Catío (CTO), 352–353
 E'ñapa Woromaipu (PBH), 352, 363
 Enets (ENH), 72
 Enga (ENQ), 266, 273
 Enggano (ENO), 269
 English (ENG), 1–3, 6, 8, 10, 16–17, 22–23, 39–40, 57, 60, 62–63, 69, 71, 80, 85–86, 104, 107, 111–112, 114, 150–151, 157, 160, 220, 222, 224, 230, 254, 270–271, 274, 277–278, 280, 298, 327, 332–333, 359, 365, 387, 401–403, 405–409, 413, 417, 423–424, 426–429, 435
 Enlhet (ENL), 352, 359
 Erzya (MYV), 72–73
 Ese Eija (ESE), 352, 357
 Eskimo, 16–17, 327–330, 363, 365, 370
 Eskimo-Aleut, 136, 149, 326–330, 344, 363
 Esperanto (EPO), 403
 Estonian (EST), 71–74
 Ethiopic, 200, 202
 Even (EVE), 128, 130
 Evenki (EVN), 128, 130
 Ewe (EWE), 213, 216
- Faliscan (XFA), 63
 Fanagalo (FNG), 409
 Faroese (FAO), 57, 60
 Fijian (FIJ), 270, 319–320
 Filipino, 268
 Filomeno Mata-Coahuilán (TLP), 345, 348
 Finnic, 72–73, 103
 Finnish (FIN), 26, 71–74, 103–120, 123–124, 131, 160, 174–175, 217, 334, 417
 Finnish Sign Language (FSE), 416–417
 Fon (FON), 213, 216–217
 Formosan, 268
 Fox (SAC), 328, 333
 Fox-Sauk-Kickapoo, 333
 French (FRA), 4, 8, 57, 64, 71, 79–80, 157, 222, 224, 327, 346, 402–403, 408–411, 413–414, 420, 424
 French Sign Language (FSL), 416, 420, 423

- Frisian, 6, 63
 Friulian (FUR), 57, 64
 Fula (FUL), 214–215
 Fur (FVR), 199, 209–211
 Fuṣḥā Arabic, 223
 Fuuta Jalon, 213
- Gabri (GAB), 199, 204
 Gagauz (GAG), 128, 130
 Gàn (GAN), 138–140
 Ganda (LUG), 213, 216
 Garifuna (CAB), 352, 355, 362
 Garo (GRT), 138, 141
 Gascon, 79
 Gastarbeiterdeutsch, 409
 Gaulish (XTG), 57–58
 Gbaya (GSO), 213–214
 Gbe, 216–217
 Ge'ez (GEZ), 199–201
 Geg Albanian (ALN), 65
 Gelao (GIQ), 138, 144
 Georgian (KAT), 75–77
 German (DEU), 4, 57, 60, 63, 80, 289, 409, 428
 Germanic, 4, 8, 56, 59–63, 70, 74, 80, 103, 402, 426
 German Sign Language (GSG), 416
 Ghanaian Sign Language (GSE), 416
 Gikuyu (KIK), 213, 216
 Gilbertese (GIL), 266–267, 270
 Gitksan (GIT), 328, 340
 Goidelic, 58
 Gondi (GON), 138, 147
 Gooniyandi (GNI), 266, 275–276
 Gothic (GOT), 60, 62, 70
 Greater Barito, 267, 269
 Greek (ELL), 56–57, 64–65, 70–71, 80, 129, 206, 224, 274
 Greek Sign Language (GSS), 416
 Gtsang, 168
 Guadeloupean Kreyol (GCF), 413
 Guahibo (GUH), 352, 354
 Guajajára (GUB), 352, 360
 Guajibóan, 354
 Guanano (GVC), 352, 355
 Guaraní (GRN), 360–361
 Guarayu (GYR), 352, 360
 Guarijío (VAR), 328, 342
 Guatemalan Sign Language (GSM), 416
 Guató (GTA), 352, 359
 Guaykuran, 357, 359
 Guianese Creole French (GCR), 413
 Gujarati (GUJ), 43, 57, 67
 Gulf, 336
 Gulf Zoquean, 348
 Gullah (GUL), 413
- Gumuz (GUK), 199, 209
 Gunwinggu (GUP), 266, 275–276
 Gunwingguan, 275–276
 Gur, 214
 Guugu Yimidhirr (KKY), 266, 275–277
 Guyanese Creole English (GYN), 413
 Gwich'in (GWI), 328, 331
- Hadiyya (HDY), 199, 205
 Hadza (HTS), 213, 220–221
 Haida (HDN), 326, 328, 330, 332
 Haiphong Sign Language (HAF), 416
 Haitian Kreyòl (HAT), 410–415
 Hai|om (HGM), 213, 220
 Hakka (HAK), 138–140
 Halkomelem (HUR), 328, 343
 Hamitic, 207–208
 Hamito-Semitic, 207–208
 Hanoi Sign Language (HAB), 416
 Harákmbut, 355
 Hausa (HAU), 9–10, 198–199, 204, 242
 Hausa Sign Language (HSL), 416, 420
 Hawaiian (HAW), 11–14, 38, 266–267, 270, 278–298, 318, 370, 415
 Hawaiian Sign Language (HPS), 416
 Hawaii Creole (HWC), 278, 413
 Haya (HAY), 213, 216, 219
 Hebrew (HEB), 31, 63, 149, 198–201, 427
 Hidatsa (HID), 328, 336
 High Germanic, 60, 63
 Highland Oaxaca Chontal (CHD), 345
 Hiligaynon (HIL), 268
 Himalayish, 141, 165
 Hindi (HIN), 15, 57, 67–68, 168
 Hiri Motu (HMO), 266–267, 274, 409, 427
 Hittite (HIT), 56–57, 69
 Hixkaryána (HIX), 352, 362–363, 394–396
 Hlai (LIC), 138, 144
 Hmong (HMN), 143, 189–190
 Hmong Daw (MWW), 138, 143
 Hmong Don (HMF), 138, 143
 Hmongic, 142–144, 143
 Hmong-Mien, 127, 142–144, 195, 270
 Hmong Njua (HNT), 138, 143
 Ho (HOC), 138, 146
 Ho Chi Minh City Sign Language (HOS), 416
 Hokan, 339, 344
 Hokkien, 140
 Honduran Sign Language (HDS), 416
 Hong Kong Sign Language (HKS), 416
 Hopi (HOP), 328, 342
 †Hua (HUC), 213, 220
 Huambisa (HUB), 352, 356
 Huastec (HUS), 345, 350
 Huasteca Nahuatl, 341

- Huautla (MAU), 345, 347
 Huave (HUV), 340, 345, 397
 Huba (HBB), 199, 204
 Huehuetla (TEE), 345, 348
 Huichol (HCH), 328, 342
 Hungarian (HUN), 26, 71–72, 74–75
 Hungarian Sign Language (HSH), 416
 Hupa (HUP), 328, 331
- Iatê (FUN), 352, 359
 Icelandic (ISL), 57, 60
 Igbo (IBO), 213, 216–218
 Ignaciano (IGN), 352, 355
 Ijo-Congo, 214
 Ijoid, 214, 218
 Ilocano (ILO), 266–268
 Indic, 67–68, 148, 166
 Indo-European, 6–7, 56–71, 74, 80, 103, 108,
 129, 135–136, 145, 149, 197, 218, 222, 229,
 244, 270, 327, 417
 Indo-Iranian, 56, 67–71, 147, 149
 Indonesian, 268–269
 Indonesian Sign Language (INL), 416
 Indo-Pacific, 265, 275
 Indo-Pakistani Sign Language (INS), 416
 Indo-Portuguese (IDB), 413
 Ingush (INH), 75–77
 Inland Northern Arawakan, 355
 Interior Salishan, 343
 Interlingua (INA), 404
 Inuinnagtun (IKT), 328, 330
 Inuit, 327, 330
 Inuktitut (IKE), 328, 330
 Iquito (IQU), 352, 356
 Iranian, 67–68, 74, 79, 103, 148
 Irántxe (IRN), 352, 360
 Irish (GLE), 15, 57–59, 427
 Irish Sign Language (ISG), 416
 Iroquoian, 326, 337–338
 Iroquois, 337
 Island Carib (CRB), 362
 Islander Creole English (ICR), 413
 Isthmus (ZAI), 345, 347
 Italian (ITA), 4, 57, 64–65, 224, 274, 289, 409
 Italian Sign Language (ISE), 416, 423
 Italic, 56, 63–64, 70
 Itelmen (ITL), 128, 136
 Itonama (ITO), 352, 360
 Iu Mien (IUM), 138, 143
 Ixil (IXL), 345, 350
 Iyo'wujwa Chorote (CRQ), 352, 359
 Izon (IJC), 213–214
- Jabutí (JBT), 352, 360
 Jabutian, 360
- Jakalteko (JAC), 345, 350
 Jalapa de Díaz (MAJ), 345, 347
 Jalnguy, 299
 Jamaican Country Sign Language (JCS), 416, 420
 Jamaican Creole (JAM), 413
 Jamaican Sign Language (JLS), 416
 Jamiltepec (MXT), 345–346
 Japanese (JPN), 2–3, 15, 23, 25, 36–38, 40–41, 49,
 128, 130, 132–134, 146, 151, 190–194, 217,
 220, 270, 380, 389
 Japanese Sign Language (JSL), 416, 423
 Japonic, 129–130, 133–136
 Jaqaru (JQR), 352, 356
 Jarrakan, 275–276
 Javanese (JAV), 266–267, 269
 Jean, 357, 359
 Jebero (JEB), 352, 356
 Jemez (TOW), 328, 341
 Jicarilla (API), 328, 331
 Jingpho (KAC), 138, 141–142
 Jingpho-Asakian, 141
 Jirrbal. *See* Dyirbal
 Jívaro, 356
 Jivaroan, 353, 356
 Jordanian Sign Language (JOS), 416
 Juang (JUN), 138, 146
 Juba Arabic (PGA), 413
 Judeo-Berber (JBE), 203–204
 Judeo-Spanish (LAD), 57, 64
 Jurúna (JUR), 352, 361
 Juxtlahuaca (VMC), 345–346
 Ju'hoan (KTZ), 213, 220, 261
- Kaapor (URB), 352, 361
 Kaapor Sign Language (UKS), 416
 Kabardian (KBD), 75–76
 Kabyle (KAB), 199, 203
 Kagate (SYW), 138, 167
 Kaiwá (KKG), 352, 360
 Kalaallisut (KAL), 328, 330, 365
 Kala Lagaw Ya (MWP), 266, 275–276
 Kalispel-Salish (FLA), 328, 343
 Kamayurá (KAY), 352, 361
 Kam-Sui, 144
 Kamwe (HIG), 199, 204
 Kanamarí (KNM), 352, 355
 Kanembu (KBL), 211
 Kannada (KAN), 138, 147
 Kanoé (KXO), 352, 360
 Kanuri (KAU), 199, 209, 211
 Kanuri-Kanembu, 211
 Kapong, 363
 Kaqchikel (CAK), 345, 350
 Kaqchikel-Tz'utujil, 350
 Karachai-Balkar (KRC), 128, 130

- Karajá (KPI), 352, 359
 Karakalpak (KAA), 128, 130
 Karelian (KRL), 72–73, 103
 Karenic, 141–142
 Karitiána (KTN), 352, 361
 Karo (ARR), 352, 361
 Kartvelian, 75–78, 136
 Karuk (KYH), 326, 328, 338, 425
 Kashaya (KJU), 326, 328
 Kashinawa (CBS), 352, 357
 Kashmiri (KAS), 57, 67
 Katab (KCG), 213, 216
 Katla (KCR), 213–214
 Katuic, 146
 Katukinan, 355
 Kawahib, 361
 Kawaiisu (XAW), 328, 342
 Kaxararí (KTX), 352, 357
 Kayabí (KYZ), 352, 360–361
 Kazakh (KAZ), 81, 128, 130
 Kenzi (XNZ), 199, 209
 Keres (KEE), 326, 344
 Keresan, 328
 Ket (KET), 128, 136–137, 332
 Kham-Hor, 167–168
 Khams (KHG), 138, 167
 Khanty (KCA), 72–73
 Khasi (KHA), 138, 146
 Khasian, 146
 Kherwari, 146
 Khmer (KHM), 138, 146
 Khmu (KJG), 138, 146
 Khmuic, 146
 Khoe, 220–221
 Khoisan, 197, 218, 220–222
 Khwe (XUU), 213, 220
 Kiamu, 241
 K'iche' (QUC), 345, 350
 K'ichean, 350
 K'ichean-Mamean, 350
 Kichwa, 373
 Kickapoo (KIT), 328, 333
 Kija (GIA), 266, 275–276
 Kiksht (WAC), 326, 328
 Kimvita, 241
 Kimwani, 241
 Kinyarwanda (KIN), 213, 216
 Kiowa (KIO), 328, 340–341
 Kiowa-Tanoan, 326, 340–341, 343
 Kituba (KTU), 413
 Kiunguja, 241
 Kiwai, 321
 Klingon (TLH), 403
 Koalib (KIB), 213–214
 Koasati (CKU), 328, 335
 Kodava (KFA), 138, 147
 Kogi (KOG), 352–353
 Kolami (KFB), 138, 147
 Komi (KOM), 72–73
 Komuz, 209
 Kongo (KON), 213, 216, 413
 Konyak (NBE), 138, 141
 Kordofanian, 212, 214
 Korean (KOR), 49, 128–130, 132–133, 135–136,
 146, 151
 Korean Sign Language (KVK), 416
 Korku (KFO), 138, 146
 Koryak (KPY), 128, 136–137
 Kota (KFE), 138, 147
 Kpelle (XPE), 213–214
 Kra, 144
 Krio (KRI), 413
 Kriol (ROP), 413
 Kru, 214
 Kui (KXU), 138, 147
 Kuki-Naga, 141
 Kulina (CUL), 352, 355
 Kumiai (DIH), 328, 338
 Kumyk (KUM), 128, 130
 Kunama (KUN), 199, 209
 Kung, 220–221
 Kung-Ekoka (KNW), 213, 220
 Kurdish (KUR), 57, 67–68
 Kurukh (KRU), 138, 147
 Kusunda (KGG), 272
 Kutenai (KUT), 326, 328
 Kuy (KDT), 138, 146
 Kwa, 214–215
 Kwakiutl (KWK), 326, 328
 †Kx'au||'ein (AUE), 213, 220
 Kyirong-Kagate, 167
 Kyrgyz (KIR), 128, 130

 Láadan (LDN), 403
 Lacandón (LAC), 345, 350
 Ladakhi (LBJ), 138, 167
 Ladakhi-Balti, 167–168
 Ladin (LLD), 57, 64
 Lahu (LHU), 138, 141
 Lahuli-Spiti, 167–168
 Lak (LBE), 75–77
 Lakota (LKT), 328, 336
 Lalana-Tepinapa (CNL), 345, 347
 Lampung, 269
 Land Dayak, 269
 Lao (LAO), 138, 144
 Lardil (LBZ), 404
 Lashi (LSI), 138, 141
 Latin (LAT), 4, 8–9, 23, 60, 63–65, 70–71,
 79–80, 198

- Latvian (LAV), 57, 65, 427
 Laz (LZZ), 75–76
 Lepcha (LEP), 138, 141
 Lese (LES), 199, 209, 211
 Lezgi (LEZ), 75–77
 Lezgian, 77
 Liberian Pidgin English (LIR), 409
 Libyan Arabic (AYL), 224
 Light Warlpiri, 403
 Lillooet (LIL), 328, 343
 Lingua Franca (PML), 409, 412
 Lisu (LIS), 138, 141
 Lithuanian (LIT), 25, 57, 65, 70, 103
 Lithuanian Sign Language (LLS), 416
 Lojban (JBO), 403
 Loke (LOY), 138, 167
 Loloish, 141
 Louisiana Creole (LOU), 413
 Low Germanic, 60
 Lowland Mixe (MIR), 345, 348
 Lowland Oaxaca Chontal (CLO), 345
 Luiseño (LUI), 328, 342
 Luo, 210
 Lushootseed (LUT), 328, 343
 Luvian (HLU), 56–57
 Lycian (XLC), 56–57
 Lyons Sign Language (LSG), 416
- Maasai (MAS), 199, 209–210
 Maba (MDE), 199, 209
 Maban, 209
 Macanese (MZS), 413
 Macedonian (MKD), 57, 65–66, 81
 Machiguenga (MCB), 352, 355
 Macro-Algonquian, 336
 Macro-Jê, 357
 Macro-Siouan, 338
 Macushi (MBC), 352, 363
 Ma'di (MHI), 199, 209, 211
 Madurese (MAD), 266–267, 269
 Mafa (MAF), 199, 204
 Maidu (NMU), 326, 328
 Mainline Panoan, 357
 Makasae (MKZ), 266, 273
 Malagasy (MLG), 213, 222, 266–267, 269
 Malay (MSA), 9–14, 149, 223, 266–267, 269, 413
 Malayalam (MAL), 138, 147–148
 Malayo-Polynesian, 11–15, 222, 268–271
 Malayo-Sumbawan, 267, 269
 Malaysian Sign Language (XML), 416
 Malecite-Passamaquoddy (PQM), 328, 333
 Maltese (MLT), 199–200, 223
 Maltese Sign Language (MDL), 416
 Malto (MJT), 138, 147
 Mam (MAM), 345, 350
- Mamean, 350
 Manchu (MNC), 128, 130, 132
 Mandaic (MID), 199–200
 Mandan (MHQ), 328, 336
 Mandarin Chinese (CMN), 132, 138–139,
 149–166, 189–190, 271, 283, 310
 Mande, 212–215, 214, 218
 Manga Kanuri (KBY), 211
 Mangbetu (MDJ), 199, 209, 211
 Maninka (EMK), 213–214
 Mansi (MNS), 72–73
 Manx (GLV), 57–58, 426
 Maori (MRI), 266–267, 270, 278, 426
 Mapudungu, 357, 359
 Mapudungun (ARN), 352, 357–360, 359
 Maquiritari (MCH), 352, 363
 Maranao (MRW), 11–14, 266–267
 Marathi (MAR), 57, 67
 Marba (MPG), 199, 204
 Margi Central (MRT), 199, 204
 Mari (CHM), 72–73, 75
 Maricopa (MRC), 328, 338
 Maritime Sign Language (NSR), 416
 Marshallese (MAH), 266–267, 270
 Martha's Vineyard Sign Language (MRE),
 420, 422
 Maru (MHX), 138, 141
 Marúbo (MZR), 352, 357
 Masa (MCN), 199, 204
 Masalit (MLS), 199, 209, 211
 Mascoyan, 357, 359
 Matacoan, 357, 359
 Matsés (MCF), 352, 357
 Maxakalí (MBL), 352, 359
 Mayan, 340, 345, 349
 Mayangna (YAN), 352–353
 Mayo (MFY), 328, 342
 Mazahua (MAZ), 345, 347
 Mazatec, 347
 Mbour Sign Language, 420
 Mbugu (MHD), 403
 Mbyá Guaraní (GUN), 352, 360
 Media Lengua (MUE), 402–403
 Mednyj Aleut (MUD), 327, 403
 Mehri (GDQ), 199–201
 Mende (MEN), 213–214
 Menominee (MEZ), 328, 333
 Meriam (ULK), 266, 273, 275
 Mescalero-Chiricahua (APM), 328, 331
 Metlatónoc (MXV), 345–346
 Mexican Sign Language (MFS), 416
 Mezquital (OTE), 345, 347
 Miahuatlán (ZAM), 345, 347
 Michif (CRG), 402–403
 Michoacán Nahuatl (NCL), 396

- Micmac (MIC), 328, 333
 Micronesian, 267–268, 270
 Middle Chinese (LTC), 17–19, 137, 139, 163–164
 Middle English (ENM), 2–3
 Middle Japanese, 2
 Midland Mixe, 348
 Mienic, 142–144, 143
 Mikasuki (MIK), 328, 335
 Mĩn, 137, 139–140, 150
 Minica Huitoto (HTO), 352, 356
 Mĩnnán, 280
 Miri (MRG), 138, 141
 Mískito (MIQ), 352–353
 Missouri River Siouan, 336
 Misumalpan, 344, 351, 353
 Miwok (SKD), 326, 328
 Mixe, 348
 Mixean, 348
 Mixe-Zoquean, 340, 345–349
 Mixtec, 346
 Miyako (MVI), 134
 Mizo (LUS), 138, 141
 Mobilian Jargon (MOD), 409
 Mocho (MHC), 345, 350
 Mocoví (MOC), 352, 359
 Modern Lhasa Tibetan, 181
 Modern South Arabian, 200, 202
 Modern Written Tibetan, 166
 Mogholi (MHJ), 128, 130
 Mohawk (MOH), 328, 337
 Mojave (MOV), 328, 338
 Moken (MWT), 266–267
 Moklen (MKM), 269
 Moksha (MDF), 72–73
 Mon (MNW), 138, 146
 Mondé (MND), 352, 361
 Mongolian (MON), 128, 130–131, 166
 Mongolian Sign Language (MSR), 416
 Mongolic, 127, 130–132, 135–136
 Monic, 146
 Mon-Khmer, 145–146
 Mono (MNR), 328, 342
 Mooré (MOS), 213–214
 Mordvinic, 72–73
 Moriori, 426
 Morisyen (MFE), 413
 Moroccan Arabic (ARY), 224
 Moru (MGD), 199, 209
 Motu (MEU), 409
 Movima (MZP), 352, 360
 Moxo, 355
 Mubi (MUB), 199, 204
 Munda, 142, 145–146, 148
 Mundari (UNR), 138, 146
 Mundurukú (MYU), 352, 361
 Murrinh-Patha (MWF), 266, 275–276
 Murui Huitoto (HUU), 352, 356
 Musey (MSE), 199, 204
 Muskogean, 326, 335–336
 Muskogee (MUS), 328, 335
 Mycenaean Greek (GMY), 64
 Na-Dene, 137, 326, 330–332, 341, 344
 Naga Pidgin (NAG), 413
 Nahuatl, 349, 396–397
 Naiki (NIT), 138, 147
 Nakh, 75, 77
 Nama (NAQ), 213, 220–221
 Nambiquaran, 360
 Nanai (GLD), 128, 130
 Nancere (NNC), 199, 204
 Nandeva (TPI), 352, 360
 Nass-Gitksan, 340
 Naukan Yupik (YNK), 328, 330
 Navajo (NAV), 25, 328, 331, 426–427
 Naxi (NXQ), 138, 141
 Nenets (YRK), 72
 Nepalese Sign Language (NSP), 416
 Nepali (NEP), 57, 67–68
 Newari (NEW), 138, 141
 New Zealand Sign Language (NZS), 416
 Nez Perce (NEZ), 328, 339
 Ngäbere (GYM), 352–353
 Nganasan (NIO), 72
 Ngbaka (NGA), 213–214
 Nheengatú (YRL), 352, 360
 Nicaraguan Sign Language (NCS), 416, 421
 Nicobarese, 146
 Niger-Congo, 197, 210–220, 241, 265
 Nigerian Pidgin English (PCM), 408–409
 Nigerian Sign Language (NSI), 416
 Nilo-Saharan, 197, 208–212
 Nilotic, 209–210
 Nisga'a (NCG), 328, 340
 Nivaclé (CAG), 352, 359
 Nivkh (NIV), 128, 136–137
 Nobiin (FIA), 199, 209–210
 NOI, 216
 Nomatsiguenga (NOT), 352, 355
 Nootka (NUK), 326, 328, 408
 Nopala (CYA), 345, 347
 North Amazonian Cariban, 363
 North Dravidian, 148
 Northeastern Aramaic, 200
 Northeastern Turkic, 130
 Northeast Zoque (ZOR), 345, 348
 Northern, 345
 Northern Alaskan Iñupiaq (ESI), 328, 330
 Northern Arawakan, 354
 Northern Barbacoan, 353

- Northern Berber, 203
 Northern Caddoan, 338
 Northern Dravidian, 147
 Northern Emberá (EMP), 352–353
 Northern M̄in (MNP), 138–139
 Northern Ngbandi (NGB), 413
 Northern Pame (PMQ), 345, 347
 Northern Sami (SME), 72–73
 Northern Sinitic, 139
 Northern Sotho (NSO), 213, 216
 Northern Tai, 144
 Northern Tepehuán (NTP), 328, 342
 Northern Tiwa (TWF), 328, 341
 Northern Tlaxiaco (XTN), 345, 347
 Northern Totonac (TOO), 346, 348
 Northern Wú, 139
 Northern Yukaghir (YKG), 128, 136
 North Ethiopic, 200
 North Germanic, 60, 62, 104
 North Highland Mixe (MTO), 345, 348
 North Levantine Arabic (APC), 224
 North Munda, 146
 North Volta-Congo, 214
 Northwestern Sinitic, 139
 Northwestern Turkic, 130
 Northwest Germanic, 4, 6, 8, 60
 Northwest Semitic, 200
 Northwest Sumatra-Barrier Islands, 269
 North Zoque (ZOS), 345, 348
 Norwegian (NOR), 25–26, 57, 60, 62, 409
 Nostratic, 136
 Nubian, 209–210
 Nuer (NUS), 199, 209–210
 Numic, 341–342
 Nupe (NUP), 213, 216–217
 Nyahkur (CBN), 138, 146
 Nyigina (NYH), 266, 275–276
 Nyishi (NYZ), 138, 141
 Nyulnyulan, 275–276
- Oaxaca Mixean, 348
 Occitan (OCI), 57, 64, 79
 Oceanian, 266
 Oceanic, 267, 270, 278
 Ocotlán (ZAC), 345, 347
 Ofayé (OPY), 352, 359
 Oirat (XAL), 128, 130–131
 Ojibwa (OJI), 328, 333
 Ojiltán (CHI), 345, 347
 Okanagan (OKA), 328, 343
 Okinawan (RYU), 128, 130, 134
 Old Cham, 269
 Old Chinese (OCH), 69
 Old Church Slavonic (CHU), 66, 71
 Old English (ANG), 2–3, 103
 Old French (FRO), 63
- Old Japanese (OJP), 2–3
 Old Javanese, 269
 Old Malay, 269
 Old Norse (NON), 60, 62–63
 Old Nubian (ONW), 199, 209
 Old Persian (PEO), 57, 67, 69
 Old Prussian (PRG), 65
 Old Tibetan (OTB), 188
 Oluta Popoluca (PLO), 345, 348
 Omaha-Ponca (OMA), 328, 336–337
 Omotic, 200, 205–206
 Oneida (ONE), 328, 337
 Ongbe (ONB), 138, 144
 Onondaga (ONO), 328, 337
 O'odham (OOD), 328, 342
 Oromo (ORM), 205
 Oscan (OSC), 63
 Ossetic (OSS), 57, 67
 Oto-Manguean, 345–346
 Otomi, 346–347
 Oto-Pamean, 347
 Oto-Pame-Chinantecan, 347
 Ottawa (OTW), 333
- Pacific Gulf Yupik (EMS), 328, 330, 364
 Pacific Northwest sprachbund, 344
 Páez (PBB), 352–353
 Paezan, 351, 353
 Pa-Hng (PHA), 138, 143
 Pai, 338
 Paipai (PPI), 328, 338
 Paiwan (PWN), 266–267
 Palaihnihan, 326
 Palantla (CPA), 345, 347
 Palau (PAU), 266–268
 Palauan, 270
 Palaungic, 146
 Palenquero (PLN), 413
 Paleosiberian, 127, 136–137, 195
 Palikúr (PLU), 352, 354
 Pama-Nyungan, 266, 275–276, 298
 Pame, 347
 Panamanian Sign Language (LSP), 416
 Panoan, 357
 Panoan Katukína (KNT), 352, 357
 Paó (BLK), 138, 141
 Papantla (TOP), 345, 348
 Papiamentu (PAP), 413
 Papuan, 265, 270, 272–276, 279, 321, 429
 Paraguayan Guaraní (GUG), 352, 360
 Paraguayan Sign Language (PYS), 416
 Parauk (PRK), 138, 146
 Parecís (PAB), 352, 354
 Pashto (PUS), 57, 67, 229
 Patamona (PBC), 352, 363
 Patwin (PWI), 326, 328

- Pawnee (PAW), 328, 338
 Peba-Yaguan, 353
 Pemon (AOC), 352, 363
 Penang Sign Language (PSG), 416
 Penutian, 340
 Peripheral Quechua, 373
 Permic, 72–73
 Persian (FAS), 45, 57, 67–68, 224, 229
 Persian Sign Language (PSC), 416
 Peruvian Sign Language (PRL), 416
 Philippine, 267–268
 Philippine Sign Language (PSP), 416
 Phoenician (PHN), 42–43, 199–201
 Piapoco (PIO), 352, 355
 Piaroa (PID), 352, 363
 Pijin (PIS), 413
 Pilagá (PLG), 352, 359
 Pima Bajo (PIA), 328, 342
 Pipil (PPL), 328, 341–342
 Pirahã (MYP), 352, 360–361
 Pisaflores (TPP), 345, 348
 Pitjantjatjara (PJT), 275–276
 Plains Cree (CRK), 403
 Plains Indian Sign Language (PSD), 416, 422
 Platoid, 216
 Polish (POL), 25, 57, 65
 Polish Sign Language (PSO), 416
 Polynesian, 267, 270, 278
 Pomoan, 326, 338
 Popolocan, 347
 Popolocan-Zapotecan, 347
 Popoluca, 348
 Poqom, 350
 Poqomchi' (POH), 345, 350
 Poqom-K'ichean, 350
 Portuguese (POR), 4, 57, 64, 222–223, 274, 402, 407, 409, 413
 Portuguese Sign Language (PSR), 416
 Potawatomi (POT), 328, 333
 Pre-Andine Arawakan, 355
 Proto-Algonquian, 333
 Proto-Anglo-Frisian, 6
 Proto-Austronesian, 271
 Proto-Dene-Yeniseian, 137
 Proto-Dravidian, 148
 Proto-East Germanic, 4, 6
 Proto-Germanic, 4, 6, 103
 Proto-Indo-European, 65–66, 74, 148
 Proto-Malayo-Polynesian, 12–13, 272
 Proto-Mixe-Zoquean, 348
 Proto-Mongolic, 131
 Proto-Niger-Congo, 212
 Proto-North Germanic, 4
 Proto-Northwest Germanic, 4, 6, 8
 Proto-Polynesian, 279
 Proto-Semitic, 201, 225
 Proto-Sino-Tibetan, 142
 Proto-Uralic, 74
 Proto-Uto-Aztecan, 341, 349
 Proto-West Germanic, 4
 Providencia Sign Language (PRZ), 416, 420
 Puerto Rican Sign Language (PSL), 416
 Puinave (PUI), 352, 355
 Puinavean, 355
 Pulaar (FUC), 213–214
 Pular (FUF), 214
 Pumé (YAE), 352, 356
 Punjabi (PAN), 57, 67
 Purik (PRX), 138, 167
 Purus Arawakan, 355
 Pütönghua, 150
 Putumayo, 373
 Pye Krumen (PYE), 213–214
 Qabiao (LAQ), 138, 144
 Q'anjob'al (KJB), 345, 350
 Q'anjob'alan, 350
 Q'anjob'alan-Chujean, 350
 Qawasqar (ALC), 352, 359
 Q'eqchi' (KER), 345, 350
 Qhichwa, 373
 Qiandong Miao (HMS), 138, 143
 Qiang, 138
 Qiangic, 141
 Quebec Sign Language (FCS), 416
 Quechan (YUM), 328, 338
 Quechua (QUE), 45, 372–394, 402–403
 Quechuan, 356
 Querétaro (OTQ), 345, 347
 Quiotepec (CHQ), 345, 347
 Ramarama, 361
 Rapa Nui (RAP), 266–267, 270
 Red Karen (KYU), 138, 141
 Rejang (REJ), 269
 Réunion Creole French (RCF), 413
 Riff, 203
 Rikbaktsa (RKB), 352, 359
 Romance, 4, 8, 63–64, 404, 426
 Romani (ROM), 57, 67–68
 Romanian (RON), 4, 57, 64
 Romansch (ROH), 57, 64
 Rotokas (ROO), 279
 Rotuman (RTM), 270
 Rundi (RUN), 213, 216
 Russenorsk, 409
 Russian (RUS), 39, 57, 65, 67, 70–71, 80–104, 108, 110–112, 115, 121–123, 132, 155, 157, 160, 172, 174, 202, 234, 244–245, 304, 327, 334, 365, 379, 388, 403, 409
 Russian Sign Language (RSL), 416
 Ryukyuan, 130, 134

- Sabanê (SAE), 352, 360
 Sahaptian, 326, 339–340
 Sahaptin, 339
 Saharan, 209–211
 Saint Lucian Creole French (ACF), 413
 Salar (SLR), 128, 130
 Sáliba (SLC), 352, 363
 Salishan, 326, 343–344
 Sálivan, 363
 Salvadoran Sign Language (ESN), 416
 Sami, 72–73
 Samoan (SMO), 11–14, 266–267, 270, 278
 Samoyedic, 72–73, 149
 San Blas Kuna (CUK), 352–353
 Sandawe (SAD), 213, 220–221
 Sango (SAG), 413
 San Jerónimo Tecóatl (MAA), 345, 347
 San Juan Colorado (MJC), 345, 347
 Sanskrit (SAN), 57, 60, 67–71, 148, 166, 169, 172
 Santali (SAT), 138, 146
 Santa María Chimalapa Zoque (ZOH), 345, 348
 Sapé (SPC), 352, 363
 Saramaccan (SRM), 413
 Sardinian (SRD), 57, 64
 Sateré-Mawé (MAV), 352, 361
 Saudi Arabian Sign Language (SDL), 416
 Sawknah (SWN), 199, 203
 Sayula Popoluca (POS), 345, 348
 Scots (SCO), 57, 63
 Scottish Gaelic (GLA), 57–58, 427
 Sechelt (SEC), 328, 343
 Seediq (TRV), 266–267
 Selkup (SEL), 72–73
 Sema (NSM), 138, 141
 Seminole, 335
 Semitic, 198–202, 200, 207–208, 223, 230, 233
 Seneca (SEE), 328, 337
 Serbian (SRP), 66, 81
 Serbo-Croatian (HBS), 25, 39, 57, 65–66
 Seri (SEI), 326, 328, 338
 Seselwa (CRS), 413
 Seward Peninsula Iñupiaq (ESK), 328, 330
 S'gaw (KSW), 138, 141
 Shambala (KSB), 213, 216, 219
 Shan (SHN), 138, 144
 Sharanahua (MCD), 352, 357
 Shawnee (SJW), 328, 333
 Sherpa (XSR), 138, 167–168
 Sherpa-Jirel, 167
 Shilha, 203
 Shipibo, 357
 Shipibo-Conibo (SHP), 352, 357
 Shona (SNA), 213, 216
 Shoshone (SHH), 328, 342
 Shuar (JIV), 352, 356
 Shuswap (SHS), 328, 343
 Sidamo (SID), 199, 205
 Sierra Popoluca (POI), 345, 348
 Sierra Totonac (TOS), 345, 348
 Signing Exact English, 417
 Sign Language of the Netherlands (DSE), 416
 Silacayoapan (MKS), 345, 347
 Singapore Sign Language (SLS), 416
 Sinhala (SIN), 57, 67–68
 Sinitic, 137–140, 142, 149
 Sino-Tibetan, 127, 137–142, 144, 149, 165, 195,
 197, 229, 271
 Siouan, 326, 336–338
 Siouan-Catawban, 336
 Slavey (XSL), 328, 331
 Slavic, 63–67, 74
 Slovak (SLK), 57, 65
 Slovakian Sign Language (SVK), 416
 Slovenian (SLV), 57, 65–66
 Somali (SOM), 198–199, 205
 Songhai (SON), 199, 209–211
 Sora (SRB), 138, 145–146
 Sorbian (WEN), 57, 65–66
 Sotho (SOT), 213, 216
 South African Sign Language (SFS), 416
 South Amazonian Cariban, 363
 South Arabian, 201
 South Bolivian Quechua (QUH), 352, 356
 South-Central Dravidian, 147
 Southeastern Turkic, 130
 Southern Arawakan, 355
 Southern Barbacoan, 353
 Southern Dene, 331
 Southern Dravidian, 147
 Southern Emberá (CMI), 352–353
 Southern Kiwai (KJD), 321
 Southern M'In (NAN), 138–140
 Southern Mongolic, 130
 Southern Nambikuára (NAB), 352, 360
 Southern Numic, 342
 Southern Qiang (QXS), 141
 Southern Quechua, 373
 Southern Sami (SMA), 72
 Southern Slavic, 65
 Southern Tepehuán (STP), 328, 342
 Southern Tiwa (TIX), 328, 341
 Southern Wú, 139
 Southern Yukaghir (YUX), 128, 136
 South Ethiopic, 200
 South Halmahera–West New Guinea, 270
 South Highland Mixe, 348
 South Midland Mixe (MXQ), 345, 348
 South Munda, 146
 South Semitic, 201
 South Slavic, 66

- South Sulawesi, 269
 South Volta-Congo, 214
 Southwestern Sinitic, 139
 Southwestern Tai, 144–145
 Southwestern Turkic, 130
 Spanish (SPA), 4, 8, 57, 64, 79, 224, 277, 333, 352,
 357, 362, 373–374, 376, 386–388, 402–403,
 409, 413
 Spanish Sign Language (SSP), 416
 Spiti Bhoti (SPT), 138, 167
 Spokane (SPO), 328, 343
 Squamish (SQU), 328, 343
 Sranan (SRN), 413
 Sri Lankan Creole Malay (SCR), 413
 Sri Lankan Sign Language (SQS), 416
 Stoney (STO), 328, 336
 Straits Salish (STR), 328, 343
 Sudanese Arabic (APD), 224
 Sui (SWI), 138, 144
 Sumerian, 35, 37–38, 41, 45–46, 48, 206
 Sundanese (SUN), 266–267, 269
 Svan (SVA), 75–76
 Swahili (SWA), 16–17, 210, 213, 216–218, 222,
 241–259, 424
 Swazi (SSW), 213, 216
 Swedish (SWE), 25–26, 57, 60, 62, 103–104
 Swedish Sign Language (SWL), 416
 Syriac (SYC), 199–201

 Taa-!Kwi, 220–221
 Tabassaran (TAB), 75–77, 110
 Tacana (TNA), 352, 357
 Tacanan, 357
 Tachelhit (SHI), 199, 203
 Tagalog (TGL), 266–268, 319
 Tahaggart Tamahaq (THV), 199, 203
 Tahitian (TAH), 266–267, 270, 278
 Tai, 144
 Tai Dam (BLT), 138, 144
 Tai-Kadai, 127, 142, 144–145, 195, 270
 Taino (TNQ), 333, 352, 354
 Taiwanese Sign Language (TSS), 416
 Tajik (TGK), 57, 67–68, 81
 Takic, 342
 Tamasheq, 203
 Tamazight (TZM), 199, 203–204
 Tamboran, 423
 Tamil (TAM), 44, 138, 147
 Tangkhul (NMF), 138, 141
 Tangut (TXG), 128, 141
 Tani, 141
 Taqbaylit, 203
 Taracahitic, 342
 Tarahumara (TAR), 328, 342
 Tarahumaran, 342
 Tarascan, 345
 Tariana (TAE), 402
 Tarifit (RIF), 199, 203
 Tasmanian, 265, 272
 Tawallammat Tamajaq (TTQ), 199, 203
 Tayart Tamajej (THZ), 199, 203
 Tebul Sign Language (TSY), 416, 420
 Teco-Mam, 350
 Tegali (RAS), 213–214
 Telugu (TEL), 138, 147
 Temiar (TEA), 138, 146
 Temoaya (OTT), 345, 347
 Tenango (OTN), 345, 347
 Tenetehara, 360
 Tenharim (PAH), 352, 361
 Tenino (TQN), 328, 339
 Tepehua, 346, 348
 Tepehuan, 341
 Tepiman, 342
 Tepo Krumen (TED), 213–214
 Tequistlatecan, 339, 345
 Terêna (TER), 352, 355
 Tetum (TET), 266–267, 269, 274
 Tewa (TEW), 328, 341
 Thado (TCZ), 138, 141
 Thai (THA), 25, 138, 144, 151
 Thai Sign Language (TSQ), 416
 Thompson (THP), 328, 343
 Tibetan, 142, 148, 188, 229
 Tibeto-Burman, 137, 140–142, 165
 Ticuna (TCA), 352, 355
 Tigré (TIG), 199–200
 Tigrinya (TIR), 199–201
 Timbuktu Tamasheq (TAQ), 199, 203
 Tiv (TIV), 213, 216, 219
 Tiwa, 341
 Tiwi (TIW), 266, 275–276
 Tlachichilco (TPT), 345, 348
 Tlacolulita (ZPK), 345, 347
 Tlapanec (TPX), 338, 345, 347
 Tlingit (TLI), 328, 330–331
 Toba (TOB), 352, 359
 Toba-Mascoy (TMF), 352, 359
 Tocharian, 56, 69–71
 Tocharian A (XTO), 56–57
 Tocharian B (TXB), 56–57, 69
 Toda (TCX), 138, 147
 Tojolabal (TOJ), 345, 350
 Tok Pisin (TPI), 409, 414, 427–445
 Tol (TIC), 338, 345
 Tolai (KSD), 428, 438
 Tolowa (TOL), 331
 Tongan (TON), 266–267, 270
 Torres Strait Creole (TCS), 413
 Tosk Albanian (ALS), 65

- Totonac, 346, 348
 Totonacan, 340, 345–346
 Trans–New Guinea, 273, 321
 Trinidadian Creole English (TRF), 413
 Trinitario (TRN), 352, 355
 Tri-State Panoan, 357
 Trukese (CHK), 266–267, 270
 Trumai (TPY), 352, 360
 Tsang, 167–168
 Tsez (DDO), 75–77
 Tsezic, 77
 Tsimané (CAS), 352, 357
 Tsimshian (TSI), 340
 Tsimshianic, 340
 Tsou (TSU), 266–267
 Tswana (TSN), 213, 216
 Tu (MJG), 128, 130
 Tuareg (TMH), 203–204
 Tübatulabal (TUB), 328, 342
 Tucano (TUO), 352, 355
 Tucanoan, 353, 355
 Tulu (TCY), 138, 147
 Tungusic, 127, 130, 132, 135–137
 Tunisian Arabic (AEB), 224
 Tunisian Sign Language (TSE), 416
 Tupari (TPR), 352, 361
 Tupí (TPW), 360
 Tupian, 360
 Tupí-Guaraní, 360
 Turkic, 79–80, 127, 129–131, 130–131,
 135–137, 229
 Turkish (TUR), 19–20, 55, 65, 124–125, 127–131,
 134, 148, 224
 Turkish Sign Language (TSM), 416
 Turkmen (TUK), 128–130
 Turoyo (TRU), 199–200
 Tuscarora (TUS), 328, 337
 Tuvan (TYV), 128, 130
 Tzeltal (TZH), 345, 350
 Tzeltalan, 350
 Tzotzil (TZO), 345, 350
 Tz'utujil (TZJ), 345, 350
- Ŭ, 167–168
 Ubykh (UBY), 75–76, 423
 Udmurt (UDM), 72–73
 Ukrainian (UKR), 57, 65, 81
 Umatilla (UMA), 328, 339
 Umbrian (XUM), 63
 Upland Yuman (YUF), 328, 338
 Ural–Altaic, 135
 Uralic, 71–75, 80, 103, 129, 131, 135–136,
 149, 417
 Urarina (URA), 352, 356
 Urdu (URD), 57, 67–68, 229
- Uru (URE), 352, 356
 Uruguayan Sign Language (UGY), 416
 Usila (CUC), 345, 347
 Ute–Chemehuevi (UTE), 328, 342
 Utian, 326
 Uto–Aztecan, 326, 341–344
 Uygur (UIG), 128, 130–131, 229
 Uzbek (UZN), 81, 128, 130
- Venezuelan Sign Language (VSL), 416
 Vietic, 146
 Vietnamese (VIE), 16, 45, 49, 138, 146, 151
 Volga Tatar (TAT), 128, 130
 Volta–Congo, 214
 Volta–Niger, 215–217
- Waiwai (WAW), 352, 363
 Wakashan, 326, 344
 Walla Walla (WAA), 328, 339
 Waorani (AUC), 352, 356
 Wapishana (WAP), 352, 354
 Warao (WBA), 352, 363
 Waray–Waray (WAR), 268
 Wari' (PAV), 352, 354
 Warlpiri (WBP), 266, 275–277, 403, 421–422
 Washo (WAS), 326, 328, 338
 Wawa, 408
 Wayampi (OYM), 352, 361
 Wayampí, 361
 Wayana (WAY), 352, 363
 Wayuu (GUC), 352, 355
 Welsh (CYM), 57–58, 70, 346, 427
 West Caucasian, 75–77, 221
 West Central Oromo (GAZ), 199, 205
 West Chadic, 204
 Western Apache (APW), 328, 331
 Western Arabic, 223
 Western Aramaic (AMW), 199–200
 Western Arrernte (ARE), 266, 275–276
 Western Austroasiatic, 145
 Western Dani (DNW), 266, 273
 Western Desert, 266, 276–277
 Western Frisian (FRY), 57, 60
 Western Highland Purepecha (PUA), 345
 Western Muskogean, 335
 Western Numic, 342
 Western Oto–Manguean, 347
 Western Popoloca (PCA), 345, 347
 Western Slavic, 65
 West Germanic, 60, 62–63
 West Scandinavian, 60
 West Semitic, 200, 202
 West Slavic, 66–67
 Wichí Lhamtés Vejoz (WLV), 352, 357, 359
 Wichita (WIC), 328, 338

- Winnebago (WIN), 328, 336
 Wintuan, 326
 Witotoan, 353, 356
 Wiyot (WIY), 332
 Wolaytta (WAL), 199, 205
 Wolof (WOL), 213–215
 Wú (WUU), 138–140

 Xavánte (XAV), 352, 359
 Xerénte (XER), 352, 359
 Xhosa (XHO), 213, 216, 218, 221
 Xiāng (HSN), 138–140
 Xiangxi Miao (MMR), 138, 143
 Xibe (SJO), 128, 130, 132
 Xinalug (KJJ), 77
 Xokleng (XOK), 352, 359

 Yagua (YAD), 352–353, 356
 Yakima (YAK), 328, 339
 Yakut (SAH), 128, 130
 Yaminahua (YAA), 352, 357
 Yanesha' (AME), 352, 354
 Yangkaal, 404
 Yanomaman, 363
 Yanomamö (GUU), 352, 363
 Yapese (YAP), 266–267, 270
 Yaqui (YAQ), 328, 342
 Yavapai, 339
 YEAL, 216
 Yeniseian, 137, 332
 Yi (III), 138, 140–142
 Yi-Burmese, 141
 Yiddish (YDD), 4, 57, 60, 63
 Yine (PIB), 352, 355
 Yokuts (YOK), 326, 328, 340

 Yokutsan, 340
 Yora (MTS), 352, 357
 Yoruba (YOR), 25, 213, 216–218
 Yosondúa (MPM), 345, 347
 Yucatec (YUA), 345, 349–350
 Yucatecan, 349–350
 Yucatecan–Core Mayan, 350
 Yucatec–Lacandon, 350
 Yucatec Maya Sign Language (MSD), 416, 420, 422
 Yuchi (YUC), 326, 328, 335, 338
 Yuè (YUE), 138–140, 150, 280
 Yukaghir, 136
 Yukpa (YUP), 352, 363
 Yuman, 326, 338–339
 Yupik, 330, 363–364
 Yupik–Inuit, 327
 Yuracare (YUZ), 352, 357
 Yurok (YUR), 328, 332–333
 Yuwana (YAU), 352, 363

 Zamucoan, 357, 359
 Zangskari (ZAU), 138, 167
 Záparo (ZRO), 352, 356
 Zaparoan, 353, 356
 Zapotec (ZAP), 347
 Zapotecan, 347
 Zempoaltepetl (MXP), 345, 348
 Zenaga (ZEN), 199, 203
 Zenzontepec (CZN), 345, 347
 Zhuang (ZYB), 138, 144
 Zoque, 348
 Zoquean, 348
 Zuberoan, 79
 Zulu (ZUL), 25, 45, 213, 216, 218, 221, 409
 Zuni (ZUN), 326, 328, 340

SUBJECT INDEX

- abessive case, 108
- abjad, 42–43, 201, 215, 225
- ablative case, 19, 110
- in Americas, 367, 378, 380
 - in Dyrirbal, 305, 308, 313
 - in Finnish, 110–111
 - in Tibetan, 176
- absolutive case, 23
- in Dyrirbal, 301–305, 310–312
 - in Tibetan, 173, 182
 - in Yup'ik, 366–367
- abstract gender, 248
- abugida, 43–44, 49–51, 132, 140, 169, 201
- accent, 25. *See also* diacritic mark; pitch accent; stress
- accusative case, 19, 23, 86, 109
- in Arabic, 228–229
 - in Europe, 86, 88, 108, 112
 - in Hawaiian, 286
 - in Quechua, 376–377
- acquisition, 419. *See also* imperfect learning; language acquisition device; nursery word
- acrolect, 415
- acrophony, 49, 226
- active voice, 92, 236
- adessive case, 110
- adjective, 97
- in Africa, 218, 243, 245, 247–248, 254
 - in Americas, 376, 380, 387–388, 402, 411
 - in Arabic, 232, 236–237
 - in Asia, 143–145, 407
 - in Australia, 277, 301–302, 306–307, 309
 - in Europe, 79–80, 88–89, 93, 111, 113–114
 - in Hawaiian, 281
 - in Sino-Tibetan, 158–159, 172, 174–176, 180
 - in Tok Pisin, 431–432, 434
- See also* comparative degree; modifier; superlative degree
- advanced tongue root, 210
- adverb, 98
- in Arabic, 229
 - in Asia, 18, 134, 162, 175, 180, 185
 - in Oceania, 314, 431–432, 439
 - in Quechua, 376, 386
 - in Russian, 84, 97–98
 - in Swahili, 257, 259
- See also* modifier
- adverbial participle, 19, 92
- adversative passive, 161
- affix, 16–22
- in Africa, 210, 218, 252, 255
 - in Afroasiatic, 202–203, 205
 - in Americas, 337, 344, 418
 - in Asia, 16, 137, 145, 148, 405
 - in Caucasus, 78
 - in creole, 414
 - in Europe, 16, 118
 - in Oceania, 273, 278, 285, 431
 - in Sino-Tibetan, 19, 142, 155, 163, 171, 177
- writing of, 38
- See also* circumfix; derivational affix; infix; inflectional affix; lexical suffix; prefix; privative affix; suffix; transfix
- affixation index, 16
- affricate consonant, 63, 151–152, 221, 277, 280, 408, 429
- agency hierarchy, 335, 381
- agent, 161, 255, 285, 287, 388
- agglutination index, 17
- agglutinative morphology, 17–22
- in Africa, 212, 217–218, 243
 - in Americas, 341, 357, 375
 - in Asia, 129, 131–133, 135, 137, 145, 148
 - in Basque, 79
 - in Caucasus, 76, 78
 - in Oceania, 273–274, 301
 - in Uralic, 74, 103, 107–108
- aggregate gender, 247–248
- agreement, 85, 91
- in Africa, 218, 243–244, 246–251, 253–255, 259
 - in Americas, 334, 337, 376, 387, 402, 411
 - in Arabic, 232, 237–238
 - in Asia, 134, 148, 158, 172
 - in Caucasus, 76, 78
 - in Dyrirbal, 301, 304, 307–309, 312–313, 315
 - in Finnish, 105, 111, 114–116
 - in Indo-European, 71, 85, 88–89, 91–92, 95
 - in Tok Pisin, 430, 432
- See also* case; gender
- AIGA, 34, 52
- alienable possession, 282–283, 290, 318
- alignment. *See* morphosyntactic alignment

- allative case, 110
 in Americas, 368, 377
 in Dyirbal, 302, 305, 308, 313
 in Tibetan, 175, 183, 185
- allograph, 226
- allomorph, 19
 in Africa, 215, 243, 253
 in Americas, 366, 391
 in Asia, 19–20, 178
 in Europe, 90, 97, 108
 in Oceania, 281, 301, 303, 306
- allophone
 in Americas, 365, 374
 in Arabic, 225
 in Europe, 82, 84, 104
 in Oceania, 277, 300, 429
 in Sino-Tibetan, 152–155, 179
 in Swahili, 243
- alphabet, 38–40
- alternative sign, 421–422
- alveolar consonant, 148, 277
- alveopalatal consonant, 152, 277, 299–300
- anagenesis, 1–3, 6
- analytic morphology, 17–22
 in Africa, 212, 217
 in American Sign Language, 418
 in Asia, 17–19, 140, 142, 145, 163
 in Finnish, 113
 in Hawaiian, 278, 281
 in Tok Pisin, 428
- ancestor, 2
- andative direction, 289
- animacy hierarchy, 334
- animate gender, 334
 in Asia, 145, 148
 in Hawaiian, 285
 in North America, 334, 340, 402
 in Russian, 88
 in Swahili, 244–247, 250–251, 255
- anterior genitive case, 302, 305
- anticipatory assimilation, 82
- antipassive voice, 307, 311–312
- apocope, 44, 168
- apophony, 59, 177, 215
- applicative voice, 251, 255
- apprehensive mood, 305, 307
- approximant consonant, 277
- Arabic script, 31, 50–51
 for Arabic, 223, 225–229
 for Berber, 204
 borrowing in, 229
 borrowing of, 45
 for Iranian, 68
 for Swahili, 241
 for Turkmen, 129
 for Urdu, 68
- Aramaic script, 50–51, 131, 225
- arbitrariness, 9–10
- archaeology, 271–272, 349
- areal feature, 79, 218
- Armenian script, 47
- article, 26, 224, 230, 237, 281–284, 304, 411
- aspect, 93, 292
 in Africa, 218, 249–250
 in Americas, 337, 380–381, 383, 411, 414, 418
 in Arabic, 234–235
 in Chinese, 158
 in creole, 414
 in Oceania, 287, 292, 307, 433
 in Russian, 92–93
See also atelicity; continuative aspect; durative aspect; frequentative aspect; habitual aspect; imperfective aspect; inchoative aspect; iterative aspect; perfective aspect; progressive aspect; punctual aspect; stative aspect; telicity
- aspirated consonant, 104
 in Africa, 221, 243, 247
 in Americas, 337, 374
 in Asia, 132, 142–143, 145, 151, 169
 in Europe, 104, 151
 in Oceania, 300, 429
- assimilation, 81, 105
 in Africa, 215, 243, 246, 251
 in Americas, 362, 390
 in Arabic, 230
 in Chinese, 154
 in Dyirbal, 300
 in Europe, 81–83, 107
- atelicity, 109–110
- ATR, 210
- augmentative gender, 246, 248
- augmentative word, 102, 291, 301
- auxiliary language, 403
- auxiliary verb, 59, 411, 414, 434
 in Americas, 335, 382
 in Arabic, 234, 238
 in Europe, 59, 113–114, 120
 in Tibetan, 177, 185–186
 in Uralic, 74
- Avestan script, 50
- avoidance register, 299
- basic vocabulary, 10
- basilect, 415
- Baybayin script, 50
- Benedictine monastery, 422
- benefactive case, 284, 377
- benefactive voice, 384
- beneficiary, 255
- Bengali script, 50
- Berber script, 47
- Beta Israel, 201

- Bible, 62, 198
 bilabial consonant, 216, 218, 404, 429
 bilingual speaker, 135, 278, 332, 373, 402
 bioprogram, 414
 borrowing, 9–10, 271, 401, 413
 in Africa, 212, 221–224
 in Americas, 327, 344, 349, 356, 365, 402, 408
 in Asia, 37, 49, 67–68, 135, 144–146, 151, 193
 in Bantu, 218, 221, 242, 246–247, 262
 in English, 39, 62–63, 224, 333, 401, 424
 in Europe, 74, 79, 103–104, 106–107, 123–124, 424
 from Americas, 333, 352
 from Arabic, 79, 224, 242, 246, 262
 from Asia, 74, 79–80, 280
 from English, 222, 224, 262, 280, 428
 from Europe, 222–224, 327, 365
 from Oceania, 223
 from Romance, 223–224, 373, 375–376, 386–387, 402
 in Indo-European, 9, 63, 65, 68, 80
 in Oceania, 272–273, 276–277, 280, 428–429, 438
 in Quechua, 373, 375–376, 386–387
 in Romance, 8, 64, 224, 333, 352
 in sign language, 417, 419
 in Sino-Tibetan, 40, 69, 140, 144, 151, 166
 in writing, 36–37, 39, 45, 51, 81, 104
 of writing, 129
See also areal feature; calque; diffusion; loanword; sprachbund
- Brahmi script, 50–51
 branch, 4
 breathy voice, 61, 63, 148
 Buddhism, 69, 165–166
 Burmese script, 51, 140
 business language, 405
- calque, 166
 Cambodian script, 51
 Canaanite script, 46, 49–51, 201
 case, 23, 85, 359, 368
 in Americas, 339, 366–370, 375–379, 386
 in Arabic, 229–232, 237–238
 in Asia, 132, 160, 172–176, 178
 in Australian, 277, 301–307, 309–310, 312–313, 315
 in Basque, 79
 in Caucasus, 77–78
 in Hawaiian, 285–287
 in Indo-European, 65–66, 71, 85–89, 92, 95–96
 in Swahili, 255
 in Uralic, 74, 107–114, 116, 118
See also abessive case; ablative case; absolutive case; accusative case; adessive case; allative case; benefactive case; comitative case; causal case; dative case; elative case; ergative case; essive case; external case; genitive case; grammatical relation; illative case; internal case; inessive case; instrumental case; intensive case; locative case; nominative case; oblique case; partitive case; perlative case; prepositional case; simulative case; terminative case; translative case; vocative case
- case particle, 173
 Catholicism, 66, 151
 causal case, 377
 causative verb, 177, 251, 384
 cause, 178, 183
 centum languages, 70
 Cham script, 50
 character, 31
 Cherokee script, 41, 47
 Chinese script, 32, 40, 47, 146, 150–151
 borrowing in, 132, 140, 151, 163–164
 borrowing of, 49
 traditional and simplified, 49, 150–151, 156
 Christianity, 64, 104, 201, 210
 circumfix, 203
 Cistercian monastery, 422
 clade, 4, 6, 268, 273
 cladistics, 44–45, 208, 272, 401, 422. *See also* comparative method; isolate
 cladogenesis, 4–6, 167, 212
 classification, 1–26, 135–136
 classifier, 137, 143–144, 157, 353, 406, 418
 semasiographic, 37, 48, 207
 clause, 310, 339, 357, 359, 361. *See also* conjunction; relative clause; verbless clause
 click, 218, 221–222, 404
 clitic, 21. *See also* enclitic; proclitic
 closed syllable, 106–107, 235, 349, 365, 429
 cluster, 406
 in Asia, 132, 143, 168–169, 171, 177
 in Oceanic, 270, 280
 in Russian, 88
 in Salishan, 343
 writing of, 43, 170–171
 coda of syllable, 41, 106, 142–143, 152, 280, 406
 code talking, 332
 cognate, 32, 135, 137, 145. *See also* cladistics
 collective number, 212
 colonialism, 222, 271, 274, 423
 color, 361
 comitative case, 379
 command, 235, 288, 306, 370
 comparative degree, 119, 121, 176, 184, 192, 232, 380
 comparative method, 9–15, 135, 333
 compounding index, 16

- compound word, 16, 155
 in American Sign Language, 418
 in Chinese, 157–160
 in Finnish, 74, 106, 118
- concept, 9–11, 197
- conditional mood, 112, 162, 178, 250, 253, 382
- conjunct consonant, 44
- conjunction, 101, 183, 256, 291, 361, 387.
See also coordinating conjunction
- conservatism, 63, 71, 167–168, 202, 325
- consonant, 142, 151–153, 221, 233
 in Africa, 207, 215, 242–243, 245–246, 251
 in Americas, 332, 337, 343, 353, 365, 374
 in Arabic, 202, 224–225, 230, 232–233, 235, 240
 in Australia, 277, 299–300, 304
 in contact language, 406, 408
 in Finnish, 104–107, 111, 113
 in Indo-European, 59, 63, 81–84, 87, 92
 in Oceania, 13, 270, 279–280, 290–291, 429
 Russian, 81
 in Tibetan, 168–169, 177, 179
 in Ubykh, 76
 writing of, 41–44, 51, 170–171, 206, 225–226, 228 (*see also* abjad; abugida; conjunct consonant; gemination; silent consonant)
See also cluster; emphatic consonant; long phoneme: consonantal; syllabic consonant; voice (phonetic)
- consonant gradation, 74, 106–108, 111, 119
- constructed international auxiliary language, 403
- constructed language, 403–404
- contact, 401, 407, 409, 424, 428
- contact language, 325, 406. *See also* creole; lingua franca; mixed language; pidgin
- content question, 94, 388
 in Finnish, 115
 in Oceania, 292, 313, 435
 in Sino-Tibetan, 161, 179
 in Yup'ik, 370
- continuative aspect, 383
- contour tone, 25, 143, 218
- contraction, 180, 235, 240
- contrafactual mood, 250
- contrast, 184, 277
- control, 78, 282, 290
- coordinating conjunction, 239
- Coptic script, 50, 210
- copula, 159, 179, 252–253
- coreferential pronoun, 369
- coronal consonant, 148, 230. *See also* alveolar consonant; dental consonant; laminal consonant; postalveolar consonant; retroflex consonant
- correspondence, 10–13, 135, 137
- creaky voice, 62, 204, 225
- Cree script, 47
- creole, 274, 409–415, 418, 427
- creole continuum, 415
- Cretan script, 47
- cuneiform script, 35, 37, 45, 47–48, 198
- cursive script, 226
- cyrillicization, 129, 131
- Cyrillic script, 45, 50–51, 226
 for East Slavic, 67
 for Mongolic, 131
 for Russian, 81
 for South Slavic, 66
 for Turkic, 129, 229
- dative case, 77, 86
 in Dyrbal, 302, 311–312
 in Georgian, 77
 in Russian, 86–87
 in Tibetan, 174–175, 182
- daughter clade, 2
- deaf-community sign, 420–421
- declarative sentence, 94, 254, 291, 370, 406
- definite reference, 26
 in Americas, 337, 367, 411
 in Arabic, 224, 228, 230–232, 237–238, 251
 in Hawaiian, 281, 283–285
 in Indo-European, 62, 96
 in Sino-Tibetan, 161, 172, 178
 in Swahili, 248
 in Tok Pisin, 440
- deletion, 8, 107, 219, 306
- deliberate action, 285, 290
- demarcative phoneme, 219
- demonstrative, 19
 in Americas, 334, 369, 402, 411
 in Asia, 144, 172, 174
 in Niger-Congo, 218
 in Oceania, 281–282, 284, 288, 304, 434, 437
See also distal demonstrative; extended demonstrative; medial demonstrative; proximal demonstrative
- Demotic Egyptian, 47
- dental consonant, 148, 277, 299. *See also* interdental consonant; labiodental consonant
- dependent, 86, 173, 229–230, 285, 359
- dependent marking, 359, 376
- derivation, 16
 in Afro-Asiatic, 205, 207
 in Arabic, 229, 232–233, 235, 239
 in Bantu, 218, 246, 248–249, 252
 in Quechua, 383, 388
 in Russian, 84, 93
 in Sino-Tibetan, 156, 177, 182
 in Tok Pisin, 433

- See also augmentative word; compound word;
 diminutive form; nominalization; patronymic;
 productiveness; transitivizer; verbalization
- derivational affix, 16
 derivation index, 16
 descendant, 2
 determiner, 114, 303
 in Americas, 337, 402, 411
 in Arabic, 236
 in Finnish, 114
 in Oceania, 157, 280–284, 303–305,
 308–309, 313
 in Swahili, 243–244, 248, 254–255, 259
 See also article
 devaluation, 396
 Devanagari script, 50, 68, 168
 devoicing, 82, 139, 429
 Dhivehi script, 50
 diacritic mark
 in Arabic, 225–226, 228–229
 in Asia, 41, 43, 132, 155, 170
 in Ge'ez, 49, 201
 in IPA, 154, 218, 300
 in Yup'ik, 365
 See also macron; marker
 dialect, 8
 dialect continuum, 66
 diffusion, 74, 268, 271, 274
 diglossia, 223, 414
 digraph, 39
 diminutive form, 18, 89–90, 151
 diminutive gender, 246, 259
 diphthong, 105, 171, 279–280, 286, 365, 429
 diptote, 230, 232
 direction, 289, 305. *See also* writing: direction of
 direct object, 286
 direct voice, 335, 358
 discourse, 282, 334
 distal demonstrative, 282, 369
 disyllabic word, 151
 documentation, 425
 double articulation, 218
 double negation, 96
 downstep, 219–220
 dual number, 69, 231
 in Americas, 340, 366–367, 369
 in Arabic, 231–232, 237–238
 in Asia, 144, 172
 in Indo-European, 71
 in Oceania, 283, 302
 dummy pronoun, 190
 durative aspect, 335, 383–384
- education. *See* oralist education; total
 communication
 Egyptian Christianity, 206
 Egyptian script, 35, 47–49, 206–207
 ejective consonant, 78
 in Afro-Asiatic, 201–202, 204–205, 207
 in Americas, 332, 337, 343, 374, 408
 in Caucasus, 78
 in Damin, 404
 elative case, 110–111, 176
 elision, 153
 ellipsis, 190–194, 309
 embedding, 361
 emphasis, 184, 252, 430
 emphatic consonant, 202, 205, 207, 224
 enclitic, 173
 in Americas, 365, 371, 378, 384–385, 388–389
 in Arabic, 233, 238–239
 in Oceania, 282, 287, 290, 313–314
 in Russian, 95
 in Tibetan, 173, 178
 endangered language, 423–427. *See also*
 extinction; language preservation; shift
 ending, 87
 in Americas, 337, 365–371, 375–381
 in Arabic, 228, 230–231
 in Chinese, 155
 in Dholuo, 210
 in Dyrbal, 301–307, 312
 in Europe, 103, 107–108, 111–113, 116
 in Russian, 87–89, 91–94, 98, 101, 121
 in Swahili, 252
 endonym, 58, 327
 ergative-absolutive alignment, 23–24
 in Americas, 357, 371
 in Asia, 71, 137, 148, 173
 in Australia, 277, 305, 310
 in Basque, 24, 79–80
 in Caucasus, 76–78
 See also split ergativity; split-S alignment
 ergative case, 21, 23
 in Basque, 79
 in Dyrbal, 301–302, 307, 310–312
 in Georgian, 78
 in Tibetan, 174, 176, 182–183
 in Yup'ik, 21, 367
 essive case, 368
 Ethiopian Orthodox Church, 201
 Ethiopic script, 50
 Etruscan script, 50
 etymology, 83
 evidentiality, 384. *See also* experiential evidence;
 hearsay evidence
 exclusive first person, 156, 283, 375, 381, 428
 existential verb, 159–160, 179–180
 exonym, 73, 203, 327
 expanded pidgin, 409, 427
 experiential evidence, 378, 381, 384
 extended demonstrative, 369

- external case, 110
 extinction, 298. *See also* genocide
- family, 7, 401
 family name, 91, 247
 feature, 133, 418
 feminine gender, 59, 244
 in Afro-Asiatic, 202–205, 207
 in Americas, 387, 402, 411
 in Arabic, 226, 229, 231, 237
 in Dyrbal, 301, 304, 308–309
 in Indo-European, 59, 85, 87, 89–90, 92
 in Niger-Congo, 218
 finger spelling, 417
 finite verb, 91–92, 112–113
 first person, 20, 85
 in Americas, 335, 370, 375, 380–382
 in Arabic, 235
 in Asia, 134, 156
 in Europe, 92, 112, 115
 in Oceania, 273, 283
 in Swahili, 249
 See also exclusive first person; inclusive first person
- focus, 319, 380, 384
 formality, 112, 414
 free variation, 279
 free word order, 79, 277, 309, 314, 334
 frequentative aspect, 291
 fricative consonant, 13
 in Africa, 216, 242
 in Afro-Asiatic, 201–202, 205, 224
 in Americas, 374
 in Asia, 134, 143, 152, 408
 in Indo-European, 63, 70
 in Oceania, 13, 277, 280, 404, 429
 in Yup'ik, 365
- front-back vowel harmony, 75, 105–106
 fusional morphology, 20–22
 in Africa, 212, 243
 in Americas, 332, 370
 in Dravidian, 148
 in Indo-European, 20, 71, 80, 87, 108
 in Uralic, 74, 108
- futhark, 62
 future tense
 in Arabic, 234
 in Europe, 92–93, 112, 120
 in Oceania, 291, 307, 432, 445
 in Quechua, 381–382
 in Sino-Tibetan, 158, 177, 180
 in Swahili, 249
- Gallaudet University, 423
 Ge'ez script, 49
- Gemara, 201
 gemination, 104
 gender, 20, 85, 87, 135
 in Africa, 212, 218, 222, 243–244, 248–251, 254–255
 in Afro-Asiatic, 202, 204–205, 207
 in Americas, 334, 369, 375, 387, 402
 in Arabic, 229, 232, 237–238
 in Asia, 135, 148, 156–157, 172
 in Finnish, 111–112
 in Indo-European, 69, 71, 85
 in Oceania, 283, 301, 304–305, 308–309, 319, 430
 in Russian, 85, 87–88, 91–92, 95
 in West Caucasian, 76
 See also abstract gender; aggregate gender; animate gender; augmentative gender; diminutive gender; feminine gender; inanimate gender; masculine gender; neuter gender; non-neuter gender; vegetable gender; vital gender
- general tense, 253
 genetic classification, 1–15, 129
 genetics, 8, 45
 genitive case, 59, 86
 in Americas, 367, 376, 379, 384
 in Arabic, 229–231, 233
 in Europe, 59, 79, 86, 106–108, 112
 in Oceania, 282–283, 301
 in Russian, 86, 96–97, 99, 101
 in Tibetan, 173–174, 180
 in Uralic, 73
 See also possession
- genocide, 423
 Georgian script, 47
 gerund, 59, 113
 gesture, 417, 419
 Glagolitic script, 47, 81
 glide, 85, 152, 235, 240
 global language, 424
 glossing, 17
 glottalic consonant, 404
 glottalization, 76, 78, 134, 221, 225
 glottalized consonant, 343
 glottal stop, 51, 62, 139, 225, 228, 280, 283
 glottogenesis, 419, 421. *See also* contact language; monogenesis; pidgin; polygenesis; sign language
- glottography, 34–44, 46
 goal, 376–377
 Gothic script, 50
 grammar, 34, 405–408, 412–414, 417, 419, 422
 grammatical relation, 23
 in Americas, 334, 339, 359, 370, 386
 in Arabic, 236

- in Asia, 19, 143, 145, 160–161, 163, 173
- in creole, 414
- in Finnish, 108
- in Indo-European, 71, 85–87, 93
- in Oceania, 285, 310, 434
- in Swahili, 254
- See also* case; dependent; direct object; head; indirect object; object; subject
- graphic communication, 32
- Greek script, 50–51, 81, 206
- Grimm's law, 60–61, 63
- Gujarati script, 43, 50
- Gurmukhi script, 50

- habitual aspect, 19, 250, 382, 418, 434
- hamza, 226, 228
- hangeul, 47, 132–133, 168
- harmony. *See* nasal harmony; vowel harmony
- head, 85
 - in Arabic, 230, 233, 236
 - in Mapudungun, 359
 - in Oceania, 274, 309
 - in Russian, 85
 - in Swahili, 248, 250, 255
- head marking, 359, 376
- hearsay evidence, 381, 384–385
- heavy nucleus, 365
- Hebrew script, 31, 50–51, 63, 204, 225
- Hieratic Egyptian, 47
- hieroglyph, 206
- Hinduism, 68, 148
- hiragana, 40, 47
- homeland. *See* urheimat
- home sign, 419–422
- homography, 36–37
- homophony
 - in Asia, 36, 43, 49, 139, 177
 - in Egyptian, 206
 - in Europe, 39, 83, 87
- homorganic consonant, 247, 302
- honorific language, 132–133, 181, 190–194
- humilific language, 134, 191

- iconic sign, 32–34, 154, 206, 417
- illative case, 110
- imperative mood, 92
 - in Americas, 370, 382
 - in Arabic, 235, 260
 - in Europe, 92, 113
 - in Oceania, 288, 306, 313
 - in Swahili, 249
 - in Tibetan, 177, 179–180, 186
- imperfective aspect, 59, 77, 93, 234–236, 287, 411
- imperfect learning, 414
- impersonal verb, 114–115
- implicational universal, 24–25
- implicit vowel, 43, 169
- implosive consonant, 204–205, 207, 212, 242
- inalienable possession, 149, 282–283, 290, 295, 315
- inanimate gender, 85, 334
- Inca, 373
- inchoative aspect, 26, 158, 160
- inclusive first person, 156, 283, 375, 381, 428
- incorporation, 21, 134, 145, 276, 337, 418
- indefinite reference, 61, 281
 - in Dyirbal, 304
 - in Quechua, 389
 - in Yup'ik, 372
- index (sign language), 418
- index (typology), 16
- indicative mood, 20, 112, 234, 252, 307, 370, 382
- indirect object, 86, 110, 174, 255, 286
- indirect speech, 435
- inessive case, 107, 110, 175
- infinitive verb, 113, 254, 411
- infix, 137
- inflection, 85, 307
 - in Africa, 212, 243, 255
 - in Afro-Asiatic, 207
 - in Americas, 365–366, 375, 380
 - in Arabic, 232–234
 - in contact language, 405–408, 411, 414
 - in Dyirbal, 301–302, 304, 306–308, 310, 312
 - in Finnish, 105, 107–115
 - in Indo-European, 66, 68–69, 71, 87–89, 91
 - in Oceania, 273, 281, 285
 - in sign language, 418, 422
 - in Sino-Tibetan, 155, 158, 173, 176, 179
- See also* diptote; ending; paradigm
- inflectional affix, 16
- inflection index, 16–17
- ingressive consonant, 404
- innateness, 414
- innovation, 2–9, 168
- instrumental case, 86
 - in Americas, 367, 379
 - in Oceania, 302, 312, 441
 - in Russian, 86–87, 97, 122
 - in Sino-Tibetan, 160, 174, 176
 - in Swahili, 255
- instrumentive voice, 307, 312
- intelligibility, mutual, 8
 - in Americas, 330, 373
 - in Basque, 8
 - in Berber, 203
 - in Sino-Tibetan, 150, 167
 - in Turkic, 131
- interdental consonant, 242
- interessive case, 378
- internal case, 110

- international language, 403
interrogative mood, 19, 370
 in Americas, 370–371, 388–389
 in Asia, 161–162, 179, 191–192
 in contact language, 406
 in Europe, 94–95, 115
 in Oceania, 291–293, 313–314, 435
 in Swahili, 254
 See also content question; question word
intonation, 95, 254, 292, 313, 418, 435.
 See also tone
intransitive verb, 21, 23–24
 in Americas, 21, 339, 366–367, 369, 372
 in Asia, 77–78, 134, 161, 173, 179, 182
 in Basque, 79
 in Oceania, 291, 306, 310–312, 433
inverse number, 341
inverse voice, 335, 358, 381
Islam, 45
 and Afro-Asiatic, 203
 and Arabic, 223–224
 and Indo-European, 68
 and Sino-Tibetan, 229
 and Turkic, 129
ISO 639-3, 55
isolate, 79–80, 134, 148, 340, 351, 360, 372
iterative aspect, 252, 307, 382
- Javanese script, 50
jer, 83, 88
Judaism, 201
jussive mood, 235, 382
- Kannada script, 50
Katakana, 47
Khmer script, 50
Koran, 45, 129, 223, 228
- labial consonant, 177, 302, 337. *See also* bilabial
 consonant
labiodental consonant, 216
Lamaism, 166
laminal consonant, 277, 299
language acquisition device, 414, 421
language preservation, 331
Lao script, 50–51
lateral consonant, 143
 in Afro-Asiatic, 202
 in Americas, 332, 408
 in Australia, 277, 404
 in Hmong-Mien, 143
latinization, 129
Latin script, 45, 50–51, 225–226
 for Berber, 204
 for Chinese, 18, 150
 for English, 39
 for Finnish, 104
 for Indo-European, 62, 66
 for Swahili, 241
 for Turkic, 81, 129, 229
 for Vietnamese, 147
 for Yup'ik, 365
lenition, 106–107
letter. *See* phonemogram
letter name, 226
level tone, 143, 218
lexical constancy, 39
lexical suffix, 344, 365
lexifier, 406, 415
ligature, 44, 226
Linear B, 47, 64
lingua franca, 198, 201
 in Africa, 204, 215, 218, 241, 408
 in Americas, 341, 361
 in Asia, 198, 201
 in Kala Lagaw Ya, 276
 in sign language, 422–423
linguistic relativity, 277, 342, 361
liquid consonant, 153, 302
loanword, 10, 271. *See also* borrowing
location
 in Arabic, 240
 in Choctaw, 335
 in Oceania, 277, 282, 284, 286, 303, 308
 in Swahili, 250, 253
 in Tibetan, 174
locational verb, 179
locative case, 24
 in Americas, 368, 377–378
 in Europe, 86, 109–111
 in Oceania, 286, 291, 301, 305, 307–308, 313
 in Sino-Tibetan, 160, 175–176
 in Swahili, 248, 255
logography, 35–38, 40, 48–49
 for Chinese, 150–151
 for Egyptian, 206
 for Greek, 64
 for Mayan, 349
logophoricity, 198, 217
long chronology, 325
long consonant, 132, 228, 364–365
long phoneme, 25, 365
 in consonant, 74
 consonantal, 104–107
 vocalic, in Finnish, 105
 in vowel in Estonian, 73
 writing of, 39, 41–43, 226, 280
long vowel
 in Arabic, 225–226, 228, 235
 in Chinese, 155
 in Hawaiian, 279–282, 286
 in Yup'ik, 365

- loss, 411
- Luvian script, 47
- macron, 280
- Malayalam script, 50
- Manchu script, 50, 132
- Mandaic script, 50
- manually coded language, 417, 420–422
- marker, 169
- masculine gender, 20, 244
 - in Afro-Asiatic, 202, 204–205, 207, 229
 - in Americas, 387, 402, 411
 - in Dyirbal, 301, 304, 308–309
 - in Indo-European, 20, 85
 - in Niger-Congo, 218
 - in Russian, 85, 89–90, 92, 99, 231, 235
- mass noun, 101, 247–248
- Maya script, 47–48, 349
- meaning. *See* semantics
- medial demonstrative, 282
- mediopassive voice, 374
- merger, 279
- mesolect, 415
- metaphor, 110
- middle voice, 381–382, 384
- minimal word, 242, 246
- mixed language, 14, 327, 401–402
- modifier
 - in Africa, 218, 243–244, 254–255
 - in Americas, 386, 411
 - in Arabic, 236–237
 - in Australia, 277, 302, 310
 - in Indo-European, 71
 - in Oceania, 274, 280–281, 291, 434
 - in Sino-Tibetan, 140, 160, 180*See also* adjective; adverb; relative clause
- monastic sign, 422
- Mongolian script, 50, 131–132
- monogenesis, 7, 412
- monosyllabic word, 40
 - in Mon-Khmer, 147
 - in Sino-Tibetan, 40, 49, 140, 142, 151, 168
 - in Tai-Kadai, 145
- Mon script, 50
- mood
 - in Americas, 369–370, 380
 - in Arabic, 234–235
 - in Finnish, 114
 - in Oceania, 287, 433
 - in Swahili, 249–252*See also* apprehensive mood; command; conditional mood; contrafactual mood; declarative sentence; imperative mood; indicative mood; interrogative mood; jussive mood; potential mood; prohibitive mood; purposive mood; sequential mood; subjunctive mood; subordinate mood
- morpheme, 16, 35–38, 107, 151. *See also* allomorph; ending; portmanteau morpheme; root; zero morpheme
- morphological typology, 16–22. *See also* agglutinative morphology; analytic morphology; fusional morphology; index (typology); polysynthetic morphology; synthetic morphology
- morphology
 - in African, 217, 222, 243–254
 - in Afro-Asiatic, 202, 205, 207
 - in Americas, 337, 365–370, 375–385
 - in Arabic, 229–236
 - in contact language, 408
 - in Europe, 22, 76–77, 85, 107–114
 - in Indo-European, 71
 - in Oceania, 280–285, 301–308, 430–434
 - in sign language, 417
 - in Sino-Tibetan, 155–162, 171
 - in Uralic, 74*See also* affix; apophony; case; consonant gradation; dependent marking; derivation; diminutive form; grammar; head marking; inflection; monosyllabic word; morpheme; paradigm; particle; reduplication; root; verb; zero morpheme
- morphophonemics, 22, 82, 173, 366, 375
- morphosyntactic alignment, 23, 334, 339. *See also* ergative-absolutive alignment; nominative-accusative alignment; split-S alignment; tripartite alignment
- mother-in-law speech, 298
- motion, 110, 289–290, 303, 376–377, 418
- multilingualism, 272. *See also* bilingual speaker
- mutation, 59, 215
- name, 107. *See also* family name; endonym; exonym; proper name
- nasal consonant
 - in Africa, 215, 221, 242–243, 246
 - in Americas, 353, 365
 - in Asia, 142–143, 153, 169, 171–172
 - in Dyirbal, 300–302*See also* prenasalized consonant
- nasal harmony, 353, 362
- nasalization, 153, 222, 337, 346
- nasalized vowel, 25, 353
- nasal spread, 353, 362
- native language, 404–405, 417, 427
- negative polarity, 19, 160
 - in Americas, 371, 388–389
 - in Europe, 96, 109, 114–115
 - in Oceania, 288, 306, 313–314
 - in Swahili, 250, 252–254, 257

- negative polarity (*Cont.*)
 in Tibetan, 180, 188
 in Uralic, 74
- neuter gender, 87, 246–247, 308–309
- neutral vowel, 124, 210
- N'ko script, 47–48, 215
- nominalization, 180, 182–183, 290–291, 371, 375, 387–388
- nominalizer, 290
- nominative-accusative alignment, 23, 339
 in Australia, 277, 305, 310
 in Caucasus, 77
 in Indo-European, 71
 in Latin, 23
 in Quechua, 376
- nominative case, 20, 23, 86
 in Arabic, 229–231
 in Finnish, 106–108, 111
 in Hawaiian, 283, 285
 in Indo-European, 86–89, 99, 103
 in Quechua, 376
- nonfinite verb, 91–92, 113, 137, 183
- nonfuture tense, 307
- non-neuter gender, 20
- nonpast tense, 92, 112
- noun, 19, 23, 93, 277, 380, 434
 in Africa, 205, 218–219, 243–249, 254–255
 in Arabic, 229–233, 237–239
 in Asia, 134, 143–145, 148–149
 in Australia, 276–277, 301–304, 308–310, 312–314
 in contact language, 402, 406–407, 411
 in Europe, 76–77, 79, 106–111, 114, 118
 in Indo-European, 65, 71
 in Munda, 145
 in North America, 334–335, 337, 340, 344, 365–369, 371
 in Oceania, 274, 280–285, 287
 in Russian, 80, 85–91, 93–95, 99, 101
 in sign language, 419
 in Sino-Tibetan, 18, 157, 161, 172–173, 181, 185
 in South America, 357, 359, 362, 374–379, 384, 386–389
See also case; definite reference; focus; gender; incorporation; indefinite reference; inflection; number; obviative noun; proximate noun; verbal noun
- number, 16
 in Africa, 218, 222, 243–244, 248–251, 254–255
 in Arabic, 229, 232, 237–238
 in Asia, 158, 172, 176, 406
 in Dyirbal, 302–303, 309
 in Finnish, 107–108, 111–114, 116
 in Indo-European, 71, 85, 87–88, 91–92, 95
 in North America, 335, 337, 340, 366, 369
 in South America, 361, 376, 387
See also collective number; dual number; inverse number; mass noun; plural number; singular number
- numeral, 144, 346, 361, 406, 418–419, 431
- nursery word, 10
- Nũshũ script, 40, 47
- object, 23–24, 26–27, 114, 380, 439
 in Arabic, 229, 233
 in Asia, 134, 148
 in Caucasus, 76–78
 in Europe, 59, 86, 93, 96–97, 101–102, 109, 112
 in North America, 334, 337, 339, 366–369, 372
 in Oceania, 273, 286–287, 289, 292, 311–312, 433–434
 in sign language, 417–418
 in Sino-Tibetan, 19, 142, 160–161, 173–174, 185–186
 in South America, 357, 359, 362, 376, 379–381, 386
 in Swahili, 249, 251, 254–255
See also direct object; indirect object
- oblique case, 229–231
- obstruent consonant, 60, 82, 429. *See also* affricate consonant; fricative consonant; Grimm's law; sibilant consonant; stop
- obviative noun, 334
- offensive language, 327
- official language, 150
- ogham, 47, 58
- Ol Chiki, 47
- Old Persian script, 47
- Old Turkic script, 129
- Old Uyghur script, 50
- onomatopoeia, 10
- onset of syllable, 106, 143, 152–153
- open syllable, 106, 235, 365, 429
- oracle bones, 151
- oralist education, 421
- oral language, 415, 417, 419, 421–423
- order. *See* word order
- ordinal number, 144
- orientation, 418
- original homeland. *See* urheimat
- Oriya script, 50
- Orkhon script, 47, 129
- Orthodox Christianity, 66, 81
- orthography, 242–243
 in Americas, 374–375, 410
 in Asia, 146, 151–155, 160, 168–171
 in Europe, 83, 104–105, 113
 in Oceania, 279–280, 300–301, 429–430
 in Yup'ik, 365
- OSV word order, 222, 314

- OVS word order, 61, 362, 403, 418
 OV word order, 371
- Pahawh Hmong, 47
 Pahlavi script, 50
 palatal consonant, 300, 302
 palatalization, 81–82, 152, 251
 palatalized consonant, 76, 87, 170, 299
 Pallava script, 50
 paradigm, 87, 176–177, 306–307
 parent clade, 2
 participle, 19
 in Americas, 370, 382–383, 387
 in Arabic, 236
 in Dyrbal, 307
 in Finnish, 113
 in Russian, 92
 See also adverbial participle
- particle, 95, 161, 172
 in Asia, 143, 158, 160–163, 172–173, 178–180, 191–192
 in Europe, 95, 115, 122
 in Oceania, 285–290, 313–314, 430, 432–433
 in Quechua, 385
 See also case particle
- partitive case, 73, 101, 109–110, 112
 part of speech, 18, 155. *See also* adjective;
 adverb; article; classifier; demonstrative;
 determiner; numeral; particle; postposition;
 preposition; pronoun; subordinator; verb
- passive voice, 92
 in Arabic, 236
 in Chinese, 161
 in Europe, 76, 92, 374
 in Finnish, 114
 in Hawaiian, 287, 290
 in Quechua, 382
 in Swahili, 251–252, 255
 See also adversative passive; antipassive voice;
 mediopassive voice
- past tense, 19
 in Americas, 355, 362, 381–382, 384, 411
 in Arabic, 234, 238
 in Europe, 92, 112–113
 in Oceania, 307, 445
 in Sino-Tibetan, 158, 165, 177, 180, 182–188
 in Swahili, 249, 254
- patient, 161
 patronymic, 91
 pausa, in, 229–232, 234
 penultimate position, 243, 280, 375
 perfective aspect, 26, 93
 in Arabic, 233–236, 238–239
 in Chinese, 158, 160
 in Georgian, 78
 in Hawaiian, 287, 292
 in Russian, 93, 112
 in Swahili, 249, 254
 in Tok Pisin, 433
- perfect tense, 113
 perlocative case, 368
 person, 380
 in Americas, 337, 358, 369, 375–376, 379–384, 387
 in Arabic, 232–233
 in Asia, 156, 158, 176, 190, 406
 in Finnish, 74, 111–114
 in Indo-European, 71, 85, 91–93
 in Oceania, 272–273, 283, 305, 310, 430–431
 in Swahili, 249, 253
 See also first person; impersonal verb; second person; third person
- phags-pa script, 168
 pharyngeal consonant, 51, 78, 205, 225, 343
 pharyngealized consonant, 202, 204, 207, 224–225
 pharyngealized vowel, 222
 Phoenician script, 42–43, 49–50, 225
 phoneme, 82, 105, 374, 417, 430. *See also*
 allophone; consonant; demarcative
 phoneme; long phoneme; tense
 phoneme; vowel
- phonemography, 38–40, 42, 105, 132, 150, 169
 phonetics
 in Americas, 365, 374–375, 417
 in Arabic, 224–225
 in Europe, 81–85, 104–105
 in Oceania, 279–281, 299–301, 429–430
 in Sino-Tibetan, 151–155, 169, 171
 in Swahili, 242–243
- phonogram, 38, 206
 phonography, 35, 38–44, 48
 phonological typology, 25–26
 phonology
 in Americas, 340, 353, 365, 374–375, 417
 in Arabic, 224–225
 in contact language, 408, 411
 in Europe, 81–85, 104–108
 in Klingon, 403
 in Oceania, 279–280, 282, 299–301, 306, 404, 429–430
 in Sino-Tibetan, 151–155, 167, 171, 406
 in Swahili, 242–243
 See also clitic; consonant gradation;
 contraction; devoicing; double negation;
 downstep; elision; feature; minimal word;
 morphophonemics; mutation; nasal
 spread; palatalization; phoneme; prosody;
 shortening; sound change; syllable
- phonotactics, 177
 pictorial symbol, 33, 37, 45, 48–49
 pidgin, 404–409, 412, 414. *See also*
 expanded pidgin

- pinyin, 18, 150–155
pitch, 25, 39, 155, 219. *See also* intonation; tone
pitch accent, 25–26, 66
place of articulation, 246
plain language, 133
plural number, 19
 in Africa, 212, 244–248, 255, 259
 in Arabic, 231–232, 235, 237, 240
 in Asia, 19, 131, 144, 156–157, 172
 in Europe, 69, 71, 74, 84–85, 87–89, 92,
 98–99, 107, 112, 115
 in North America, 340, 344, 366–367, 369
 in Oceania, 280–281, 283, 301–302, 305, 309
 in South America, 361, 375–376, 379,
 381–382, 387
 in Trans-New Guinea, 273
poetry, 162–164
polarity. *See* negative polarity
polar question, 95
 in Americas, 370–371, 388, 418
 in Europe, 95, 115
 in Oceania, 292, 313, 435
 in Sino-Tibetan, 161, 179
polite language, 133, 156
polygenesis, 7, 47
polysynthetic morphology, 17, 21–22
 in Asia, 134, 145
 in North America, 21–22, 332, 344, 365, 402,
 408, 418
 in South America, 353, 375
 in Tiwi, 276
portmanteau morpheme, 20, 22
possession, 19
 in Arabic, 229–230, 233, 237, 239
 in Asia, 19, 124, 134, 143, 149
 in Europe, 86, 97, 107, 110–111
 in Oceania, 281–284, 292, 295, 302
 in Sino-Tibetan, 159, 162, 174, 180
 in South America, 359, 375–376, 384
 in Swahili, 248–249, 257
 in Yup'ik, 366–367, 369, 371
 See also alienable possession; genitive case;
 inalienable possession
postalveolar consonant, 429. *See also*
 alveopalatal consonant; retroflex
 consonant
postposition, 18, 24, 135
 in Asia, 18, 135, 145, 156
 in Papuan, 274
 in Quechua, 386
posture, 335
potential mood, 113
prefix, 16, 135, 285, 289
 in Africa, 205, 243–251, 253–255, 257, 259
 in America, 16, 337, 362
 in Arabic, 230, 235
 in Asia, 134–135, 137, 147–149
 in Oceania, 273, 285, 301
 in Tibetan, 168, 171, 177
 in West Caucasian, 76
prefixation index, 16
prenasalized consonant, 143, 169
 in Africa, 211, 215, 242, 246
 in Guaraní, 362
 in Tibetan, 169
 in Tok Pisin, 429
preposition, 24, 114
 in Arabic, 229, 233, 239–240
 in Asia, 135, 160
 in Indo-European, 71, 86–87, 97, 99, 411
 in Oceania, 274, 431, 434, 439
 in Quechua, 374, 376, 386
 in Swahili, 255
prepositional case, 27
present tense, 20
 in Arabic, 234
 in Europe, 91–94, 109, 112
 in Oceania, 288, 307
 in Quechua, 380–382, 391
 in Swahili, 249, 253
 in Tibetan, 177, 180, 185–186, 188
prestige, 145, 151
privative affix, 180
proclitic, 230, 232, 239, 281–283, 285–289, 291
productiveness, 108
progressive aspect, 20, 186, 288, 380, 384, 433
prohibition, 180, 235
prohibitive mood, 288, 306, 313, 389
pronominalized language, 142
pronoun, 59, 190–194, 273, 369, 430
 in Africa, 198, 217, 249–255
 in Americas, 369, 375–376, 379–381, 389, 418
 in Arabic, 232–233, 238–239
 in Asia, 134, 143–144, 156–157, 172, 190, 406
 in Australia, 277, 305, 309–310, 313–315
 in Finnish, 112, 115–116
 in Indo-European, 59, 62, 85–86, 95, 99
 in Oceania, 272, 276, 283–286
 in Tok Pisin, 428, 430–431, 434, 437, 440
 See also coreferential pronoun; dummy
 pronoun; logophoricity; person; relative
 pronoun; resumptive pronoun
pronunciation. *See* phonology
proper name, 285–286
prosody, 353
protolanguage, 4–7, 271
proximal demonstrative, 19, 282, 369
proximate noun, 334
psycholinguistics, 277
punctual aspect, 445
punctuation, 171
purpose, 182, 234, 256, 285, 288, 377

- purposive mood, 307, 311
 pyromancy, 151
- quadrangular vowel system, 105
 quantifier, 361
 question. *See* interrogative mood; polar question
 question word, 313, 388, 402, 435
 quotation, 184
 Qur'an. *See* Koran
- reanalysis, 246, 290
 recent past, 362
 recipient, 255
 reciprocal voice, 252, 384
 reconstruction, 6, 13, 103, 135, 209, 271
 recurrent correspondence, 10–11, 60
 recursion, 361
 reduplicant, 301
 reduplication, 252, 291, 301–302, 309, 344
 reflexive voice, 98, 289, 307, 312, 430
 register, 133, 298–299. *See also* formality;
 honorific language; humilific language;
 mother-in-law speech; offensive language;
 official language; plain language; polite
 language; vernacular language
 relatedness, 6, 9–13
 relative clause, 95, 186, 402
 in Americas, 371, 388
 in Arabic, 238
 in Europe, 95–96, 116
 in Oceania, 292, 307, 312–313, 434–435
 in Sino-Tibetan, 160, 162, 173, 180–182, 185
 in Swahili, 249–250, 252–255, 259
 See also resumptive pronoun
 relative pronoun, 95, 116, 238, 255, 402, 435
 relexification, 407, 412
 renewal, 423–427
 resultative construction, 159–160
 resumptive pronoun, 238
 retention, 3, 6
 retroflex consonant, 148, 277, 299
 revitalization, 424–426
 rhotic consonant, 204, 277, 300, 429
 rhyme, 163
 Roman alphabet. *See* Latin script
 root, 16, 21, 74
 in Americas, 344, 349, 353, 365, 371, 402
 in Arabic, 202, 224, 232–234, 238–241, 246
 in Asia, 19, 151, 153
 in Dyrbal, 302
 in Europe, 16, 74, 93, 106–107
 in Swahili, 246–247, 249, 252
 in Tibetan, 137, 168
 writing of, 349
 Rosetta Stone, 206
 roundedness harmony, 75
 rounded vowel, 411
 runic script, 50, 61–62
- Samaritan script, 50
 sandhi. *See* tone sandhi
 satem languages, 70, 81
 school sign, 420–421
 script, 44–51, 129
 second person, 85
 in Americas, 335, 370, 375, 380–381
 in Arabic, 232, 234–235
 in Chinese, 156
 in Finnish, 112–113
 in Oceania, 273, 283, 288, 313
 in Swahili, 249, 251
 See also inclusive first person
 semantic role. *See* agent; goal; location; modifier;
 motion; purpose; patient; possession;
 quantifier; recipient; source
 semantics, 282, 308–314, 417
 semasiography, 34–35, 46, 48
 semivowel. *See* glide
 sequentiality, 31
 sequential mood, 250, 307, 311
 serial verb, 160, 217
 shift, 298, 424
 short chronology, 325
 shortening, 106
 sibilant consonant, 70
 sign, 32, 415, 419
 sign language, 415–423
 SignWriting, 47
 silent consonant, 105, 107, 113–114
 similarity, 265, 344
 simplification, 407–408
 simulative case, 285
 singular number, 20
 in Africa, 212, 244–250, 255
 in Arabic, 229, 231–232, 234–235, 237
 in Asia, 134, 156–157, 172
 in Finnish, 103, 106–108, 113, 115–116
 in Indo-European, 69, 71, 85, 87–89, 91–92
 in North America, 334–335, 340, 366, 369
 in Oceania, 273, 281, 283, 285
 in South America, 361, 381
 Sinhala script, 50
 sister clade, 4
 social status, 181
 sociolinguistics, 298. *See also* acquisition;
 business language; contact language;
 dialect; diglossia; discourse; endangered
 language; global language; glottogenesis;
 language preservation; lexifier; native
 language; prestige; register; renewal;
 revitalization; social status
 Sogdian script, 50

- sonorant phoneme, 172. *See also* approximant
consonant; glide; liquid consonant; nasal
consonant; tap; trill; vowel
- sound change, 3–6, 11–13, 271, 333, 373.
See also apocope; assimilation; merger;
tonogenesis
- source, 174, 176
- South Arabian script, 50
- SOV word order, 15, 23–24, 135
in Africa, 202, 205, 212, 218
in Altaic, 129, 131–132
in Americas, 337, 353, 357, 386
in Caucasus, 76–77
in East Asia, 23, 132–134, 180, 191
in Europe, 74, 79
in Oceania, 270, 274, 277, 314
in Paleosiberian, 137
in sign language, 419
in South Asia, 71, 145, 148
- space, 418, 421. *See also* location; motion
- speech. *See* indirect speech; quotation
- spelling reform, 411
- split ergativity, 78, 277, 305
- split-S alignment, 78–79
- sprachbund, 79, 344
- state, 287, 289
- stative aspect, 252, 281
- stop, 202, 300
in Africa, 218, 242–243, 246
in Afro-Asiatic, 207
in Americas, 337, 362, 373–374
in Asia, 132, 143, 145, 148, 408
in Indo-European, 63, 69–70
in Oceania, 277, 299–302, 429
in Sino-Tibetan, 139, 151–152, 170–171, 177
in Uralic, 74, 104, 106–107
See also glottal stop
- stress, 25–26
in Americas, 362, 365, 375, 385, 411
in Chinese, 155
in Finnish, 105
in Indo-European, 62, 84, 87–88, 92, 99, 151
in Oceania, 280, 300
in Swahili, 243
writing of, 39
- style, 93
- subgrouping, 13–15, 268
- subject, 23–24, 26, 91, 114, 339
in Afro-Asiatic, 207, 229, 236–238
in Americas, 334–335, 337, 339, 357, 359, 362
in Asia, 19, 134, 142, 148, 159–160, 173–174,
176–177, 182
in Caucasus, 76–78
in Europe, 71, 79, 86, 112, 114–116
in Oceania, 285, 288–290, 292, 310–312, 314,
430, 432, 434
in Quechua, 375, 379–383, 386–387
in Russian, 86, 91–93, 95–96
in sign language, 418
in Swahili, 244, 249–251, 253–255, 257
in Yup'ik, 366–367, 369, 371
- subjunctive mood, 97, 234, 252–253, 256,
259–260, 288
- subordinate clause, 95
in Americas, 370, 383, 387–388
in Arabic, 234
in Europe, 95–96, 115–116
in Oceania, 291–292, 312
in Sino-Tibetan, 162, 178, 183, 186, 188
in Swahili, 255–256, 259
- subordinate mood, 370
- subordinator, 97
- substrate, 406, 413
- suffix, 107, 135, 334
in Africa, 205, 215
in Arabic, 226, 228–232, 235
in Asia, 20, 131, 134, 148, 407
in Australia, 301–303, 305–307, 311–312
in Caucasus, 76–77, 110
in Europe, 20, 62, 71, 86
in Finnish, 74, 105–107, 110–114
in North America, 334–335, 337, 339, 341, 344
in Oceania, 274, 283, 290–291, 431, 433
in Quechua, 374–376, 379–384, 386–388,
391, 402
in Russian, 82–83, 85, 87, 89–90
in Sino-Tibetan, 151, 153, 157, 160, 173, 177,
180, 182–183
in South America, 353, 358, 402
in Swahili, 243, 248–249, 251–252, 255
in Uralic, 74–75
in Yup'ik, 367–369, 371
- suffix index, 16
- superlative degree, 117, 232
- superstrate, 406
- suppletion, 22, 121, 304, 335
- SVO word order, 23, 135
in Africa, 205, 212, 218, 222, 254
in Americas, 371, 386, 411, 414
in Asia, 77, 143, 145, 147, 270, 407
in Europe, 23, 61, 71, 80, 93, 103
in sign language, 417–418
in Sino-Tibetan, 140, 142, 160
in Tok Pisin, 428, 434
- switch reference, 339, 383
- syllabic consonant, 134, 242, 245–246
- syllable, 25, 151, 154–155, 163, 169
in Africa, 219, 242–243, 246
in Arabic, 235
in Asia, 134, 142, 151–153, 406
in Europe, 84–85, 105–107, 151
in North America, 343, 365

- in Oceania, 279–280, 282, 300–301, 303
 in South America, 168, 171, 353, 362,
 374–375, 385
 in Uralic, 74
 writing of, 40–42, 44, 133, 142, 169–171
See also closed syllable; coda of syllable;
 disyllabic word; heavy nucleus; onset of
 syllable; open syllable
 syllabography, 40–42, 49, 64, 140, 349, 365
 symbol, 31–44, 417
 syntactic typology, 22–25
 syntax, 24, 402
 in Africa, 198, 254–256
 in Americas, 370–371, 386–389
 in Arabic, 236–238
 in Asia, 155–162, 171, 173, 406
 in contact language, 411
 in creole, 413
 in Europe, 93–96, 114–116
 in Oceania, 277, 280–293, 308–314, 434–435
 in sign language, 417
See also agency hierarchy; agreement;
 direction; ellipsis; embedding; grammar;
 grammatical relation; incorporation;
 morphosyntactic alignment; number; part
 of speech; person; switch reference; topic;
 voice (syntactic)
 synthesis index, 16
 synthetic morphology, 17–22
 Syriac script, 50

 taboo, 298
 Tamil script, 50
 tap, 277, 429
 telicity, 109, 112
 Telugu script, 50
 tense, 16, 91
 in Americas, 342, 380–381, 411, 414
 in Arabic, 234–235
 in Asia, 158, 179, 406
 in Bantu, 218
 in creole, 414
 in Europe, 95, 112–114
 in Indo-European, 71
 in Oceania, 287, 292, 432
 in Swahili, 249–250, 254
See also future tense; general tense; nonfuture
 tense; nonpast tense; past tense; perfect
 tense; present tense; recent past
 tense phoneme, 132, 429
 terminative case, 376
 Thai script, 50–51
 third person, 19, 85
 in Arabic, 232, 235
 in Chinese, 156
 in Europe, 91–92, 99, 111, 115–116
 in North America, 335, 369–370
 in Oceania, 283, 285, 305
 in South America, 358, 360, 362, 375, 380–382
 in Swahili, 251
 in Tok Pisin, 430, 432, 437, 440
 Tibetan script, 50–51, 165–166, 168–171
 Tifinagh script, 47, 204
 time, 178, 186
 tone, 25–26
 in Africa, 198, 205, 212, 218–220
 in Americas, 332, 341, 346
 in Asia, 137, 143, 145, 147–148
 in North Germanic, 62
 in Sino-Tibetan, 140, 142, 154–155, 160, 163,
 168, 171
 See also contour tone; intonation
 tone letter, 154, 218
 tone sandhi, 155
 tonogenesis, 168
 topic, 159–160, 184, 310, 334, 358, 384, 417
 topicalization, 93, 417–418
 topic chain, 310–312
 total communication, 423
 trade, 407
 transfix, 202, 204–205, 207, 232–233
 transitive verb, 21–24
 in Americas, 334, 339, 357, 380
 in Asia, 77–78, 134, 173–174, 177, 428, 433
 in Basque, 79
 in Oceania, 285, 291, 303, 306, 310–312
 in Yup'ik, 21–22, 366–367, 369, 372
 transitivizer, 290–291
 translative case, 108
 tree, 4
 triangular vowel system, 277, 365
 trill, 211, 277, 404, 429
 tripartite alignment, 339
 typology, 1, 15–26, 265, 270, 277, 362

 Ugaritic script, 47
 underlying form, 83–84
 unergative verb, 78–79
 uniformitarian principle, 3
 United States Marine Corps, 332
 universal. *See* implicational universal
 universal language, 403
 uppercase, 226
 upward direction, 290
 urban sign, 420–421
 urheimat, 271, 373
 uvular consonant
 in Americas, 343, 365, 373–374, 408
 in Arabic, 224–225
 in Asia, 143, 148
 in Klingon, 403
 in Proto-Semitic, 202

- Vai script, 47
- valency. *See* impersonal verb; intransitive verb; transitive verb
- variation, 224. *See also* free variation
- Veda, 68–69
- vegetable gender, 308
- velar consonant
- in Africa, 218, 242
 - in Arabic, 224, 230
 - in Chinese, 152
 - in Indo-European, 69–70, 81, 84
 - in Klingon, 403
 - in North America, 343, 410
 - in Oceania, 300, 404, 429
- velarization, 84
- venitive direction, 289
- verb, 158, 183, 313–314, 337, 418
- in Africa, 217–218, 222, 243–244, 248–256
 - in Arabic, 229, 233–236, 238, 240
 - in Asia, 133–135, 137, 142, 144–145, 148, 193, 406
 - in Australia, 276, 303, 306–308, 315
 - in Caucasus, 78
 - in Chinese, 18, 156, 158–161
 - in Europe, 76–77, 79, 112–116, 118
 - in Indo-European, 59, 71, 86, 91–93, 407
 - in North America, 334–335, 337, 339, 402, 411, 414
 - in Oceania, 273, 281, 285–293, 428, 430, 432–434, 438
 - in Quechua, 375–376, 379–389
 - in sign language, 417–418
 - in South America, 353, 357, 359, 362
 - in Tibetan, 166, 174, 176–186, 188
 - in Uralic, 74
 - in Yup'ik, 365–367, 369–371
- See also* aspect; auxiliary verb; clause; copula; impersonal verb; infinitive verb; existential verb; finite verb; gerund; locational verb; nominalization; nonfinite verb; participle; serial verb; tense; transitive verb; unergative verb
- verbalization, 303, 383
- verbal noun, 239
- verbless clause, 94, 236, 281, 284–285, 309
- vernacular language, 166, 223
- vigesimal numeral, 346
- village sign, 419–422
- visibility, 303
- vital gender, 245–247
- vocabulary, 10–11, 271, 298, 408
- in Asia, 37, 135, 139, 157
 - in contact language, 405–408, 412, 414
 - in Finnish, 103
 - in Indo-European, 64, 67–68
 - in Interlingua, 404
 - in Klingon, 403
 - in Nilo-Saharan, 211
 - in Oceania, 274, 299, 404, 428
 - in sign language, 417, 420–422
 - in South America, 356, 374
- See also* color; endonym; exonym; family name; homophony; underlying form
- vocative case, 287
- voice (phonetic), 106
- in Africa, 207, 211, 221, 242, 246–247
 - in Americas, 337, 353, 362, 365, 374
 - in Arabic, 225
 - in Asia, 132, 139–140, 142–143, 145, 151, 168–169, 171
 - in Australia, 277
 - in Finnish, 104, 106
 - in Indo-European, 63, 70, 82
 - in Oceania, 300, 429
 - in Uralic, 74
- writing of, 41
- See also* aspirated consonant; breathy voice; creaky voice; Grimm's law; voice assimilation
- voice (syntactic), 76, 92, 113, 307, 383. *See also* active voice; antipassive voice; applicative voice; benefactive voice; direct voice; instrumental voice; inverse voice; mediopassive voice; middle voice; passive voice; reciprocal voice; reflexive voice
- voice assimilation, 82
- VOS word order, 222, 270
- vowel, 25, 76
- in Africa, 210, 222
 - in Americas, 353
 - in Arabic, 225, 233–236, 240
 - in Australia, 277
 - in Basque, 79
 - in Chinese, 152–154
 - in Dyrbal, 300–301, 305
 - in Finnish, 105–106
 - in Haitian Kreyòl, 411
 - in Hawaiian, 279–282, 284, 286
 - in Indo-European, 61, 66
 - in Mon-Khmer, 147
 - in North America, 337, 346
 - in Quechua, 374–375
 - in Russian, 81–84, 87–88
 - in Swahili, 243, 245–247, 250, 253–254
 - in Tibetan, 171–172, 177, 179
 - in Tok Pisin, 429–430
 - in Uralic, 74
 - in West Caucasian, 76

- writing of, 39, 42–44, 49, 51, 207, 226, 228, 349 (*see also* abugida; abjad)
- in Yup'ik, 365
- See also* diphthong; jer; long vowel; nasalized vowel; pharyngealized vowel; quadrangular vowel system; rounded vowel; triangular vowel system
- vowel harmony, 74, 135
- in Altaic, 124–125, 129, 131–132, 135
- in Asia, 132–133, 137
- in Dholuo, 210
- in Finnish, 105–106, 108
- in Uralic, 74
- See also* front-back vowel harmony; roundedness harmony
- VSO word order, 15, 23–24
- in Africa, 202, 204, 207, 212
- in Arabic, 236
- in Celtic, 59, 71
- in Láadan, 403
- in Oto-Manguean, 346
- in Polynesian, 270, 285
- Whorfian hypothesis, 277, 342–343
- word order, 15, 22–25
- in Africa, 212, 218, 222, 244, 254
- in Afro-Asiatic, 202, 204–205, 207, 236–237
- in Americas, 334, 337, 346, 357, 362, 371, 386
- in Asia, 135, 137, 145, 147–148
- in Caucasus, 76–77
- in creole, 411, 414
- in East Asia, 131–132, 134, 143, 406–407
- in Europe, 79, 103, 131
- in Indo-European, 71, 80, 93–95
- in Klingon, 403
- in Oceania, 222, 270, 274, 277, 280, 285, 309, 314, 428, 434
- in sign language, 417, 419
- in Sino-Tibetan, 18, 140, 142, 160–161, 163, 174, 179–180
- See also* free word order; OSV word order; OVS word order; OV word order; SOV word order; SVO word order; VOS word order; VSO word order
- writing, 31–51
- of Asia, 129, 131–132, 137, 140
- of Australia, 426
- direction of, 31, 42, 131–132, 226
- of Indonesia, 269
- of Maya, 48, 344, 349
- of Slavic, 66, 81
- typology of, 31–44
- Yi script, 47, 140, 142
- zero, 11
- zero morpheme, 20, 253
- Zoroastrianism, 69

