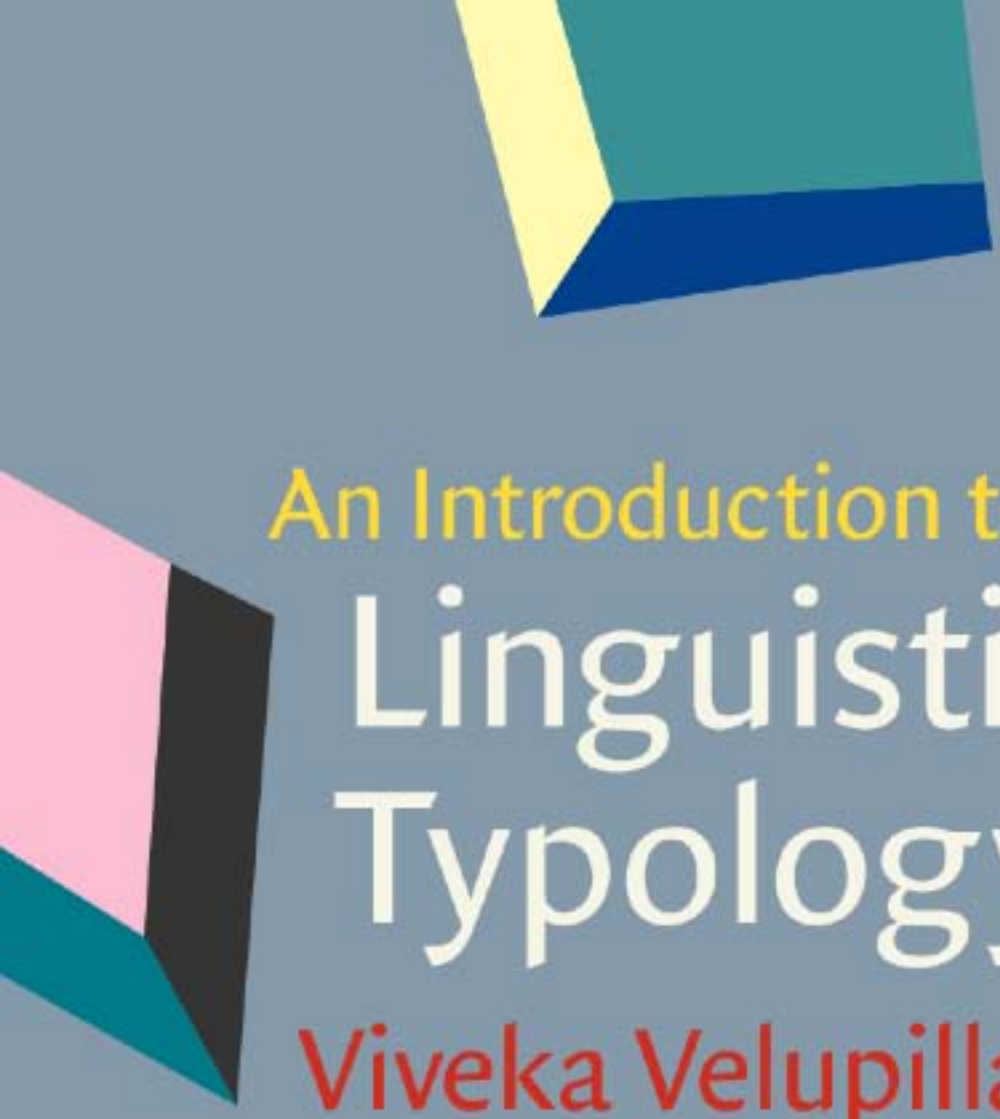


An Introduction to Linguistic Typology

Viveka Velupillai

John Benjamins Publishing Company

An Introduction to Linguistic Typology

An Introduction to Linguistic Typology

Viveka Velupillai
University of Giessen

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

An introduction to linguistic typology / Viveka Velupillai.
p. cm.
Includes bibliographical references and index.
1. Typology (Linguistics) 2. Linguistic universals. I. Title.
P204.V45 2012
415--dc23 2012020909
ISBN 978 90 272 1198 9 (Hb; alk. paper)
ISBN 978 90 272 1199 6 (Pb; alk. paper)
ISBN 978 90 272 7350 5 (Eb)

John Benjamins Publishing Company • P.O. Box 36224 • 1020 ME Amsterdam • The Netherlands
John Benjamins North America • P.O. Box 27519 • Philadelphia PA 19118-0519 • USA

For அப்பா Vela
anchor and inspiration

Table of contents

Acknowledgements

I am greatly indebted to many friends and colleagues for lending me their support during the writing of this book. In alphabetical order I wish to thank Roger Blench, Cecil Brown, Charlotta Busing, Eva Csató, Gunilla Gren-Eklund, Adam Hatfield, Lars Johanson, Kathryn Khairi-Taraki, Sven Langbein, Jeanette Sakel, Wolfgang Schulze and Bo Utas for data, comments and general encouragement. I am also grateful to Elizabeth Parks for guidance and input on sign languages and Ulrike Zeshan for letting me use her images of Lebanese Sign Language negation. Many thanks indeed to old friend from Ancient Greek days Harald Hammarström not only for very valuable comments but also for his generous donation of data, without which my pilot studies in this book would have been considerably thinner. I still owe Søren Wichmann the promised whisky, especially for his help with language locations.

Special thanks to Elly van Gelderen, Spike Gildea, Anthony Grant, Frank Lichtenberk and Tim Thornes for their generous and helpful comments on the manuscript of the book.

Special thanks also to Magnus Nissel for rushing to my rescue during hardware crashes and other such adventures.

I am extremely grateful to Maria Johanson and Patrick Thornhill for lending me their artistic skills for the illustrations, as well as to Bianca Widlitzki for her proofreading help.

Very special thanks go to Sandra Götz for her never-ending patience during my writing process.

My greatest thanks go to my husband, Magnus Huber, without whom the journey that led to this book would not have been possible.

Abbreviations

-	boundary between host and affix
~	boundary between reduplicated element and stem
=	boundary between host and clitic
<>	boundary between host and infix
*	ungrammatical expression / reconstructed form
+	elements form compound or derivative stem
//	boundary marker (i.e. pause, eye-blinks, change of facial expression)
1, 2, 3	1st, 2nd, 3rd person
I, II, III,…	noun/verb (stem) class
A	agent or the A argument of a clause
ABL	ablative
ABS	absolutive
AC	affirmative copula
ACC	accusative
ACCD	accusative definite
ACT	actual marker (on verbs)
ACTN	action
ACTV	active voice marker
ADD	additive connective
ADVZ	adverbializer
AF	agent focus
AG	agentive
AGR	agreement
ALL	allative
ANT	anterior
ANTIPASS	antipassive
AOR	aorist
APPREH	apprehensive
ART	article
ASP	aspect marker
ASS	assertive
ATT	attributive
AV	actor voice
AVOID	avoidance affix
BEG	begun aspect
BEN	benefactive
BPX	bring prefix
C	common gender
CAUS	causative
CERT	certainty modal
CF	contrastive focus
CLAN	classifier for animates
CM	class marker
CNTR	contrastive
COMP	complementizer
COMPAR	comparative case
COMPL	completive
COND	conditional
CONJ	conjunction
CONST	construct suffix
CONT	continuative
CONTR	counterfactual
COP	copula
CS	construct state
CSM	change of state marker
CTY	certainty
DAT	dative
DDET	distance determiner
DECL	declarative

DEF	definite
DEFOC	definite object conjugation
DEM	demonstrative
DEMDIST	distal demonstrative
DET	determiner
DIM	diminutive
DIR	direct
DIRC	directional
DISLOC	dislocative
DIST	distant
DP	dynamic perfect particle
DS	different subject marker
DU	dual
DUB	dubitative
EFF	effector
EMPH	emphatic
ERG	ergative
EV	evidentiality
EXCL	exclusive
EXP	experiential
EXPECT	expectational
F	feminine gender
FM	formal
FNS	final nominal suffix
FOC	focus
FRP	feminine realis past
FUT	future
FV	final vowel
GEN	genitive
GER	gerund
HAB	habitual
HPL	human plural
HS	hearsay
HUM	human
IMP	imperative
INCL	inclusive
INCOMPL	incomplete
IND	indicative
INDEF	indefinite
INDEFOC	indefinite object conjugation
INF	infinitive
INFER	inferential
INFL	inflection mode
INFR	inference
INST	instrumental
INT	intensive
INTER	interessive
INTR	intransitive
INTS	intensifier
IPF	imperfect
IPFV	imperfective
IR	irrealis
IVNPFS	intransitive verb nonperfect
	final suffix
LK	linking element
LOC	locative
M	masculine gender
ME	multiple eventhood
MOD	modal
MODIF	modifier
MS	marked scenario
N	neuter gender
N1P	non-first person plural
NCL	noun classifier
NEG	negation
NEGPOT	negative potential
NEU	neutral tense
NF	nonfeminine
NFUT	nonfuture
NH	non-honorific
NHPL	non-human plural
NHUM	non-human
NHYP	nonhypothetical
NOM	nominative
NONVIS	non-visual
NPL	non-plural
NPST	nonpast
NPT	nonpreterit
NPU	non-punctual aspect
NSG	non-singular
NZR	nominalizer
OBJ	object marker

OBL	oblique
OR	orientation prefix
ORF	object referent in focus
P	patient or the P argument of a clause
P	person
PAT	patient
PASS	passive
PB	plural bound subject pronoun
PC	present continuous/progressive
PERS	personal
PFCT	perfect
PFV	perfective
PFX	preverbal prefix
PHRF	phrase final
PL	plural
PM	plural masculine
PN	personal name
POSS	possessive
POSSR	possessor (kinsman)
POT	potential
PP	plural possessive pronoun
PPD	postpositive determinant
PPERF	past perfective
PPOS	general postposition
PREP	preposition
PRES	present tense
PRET	preterit
PREV	preverb
PRIV	privative
PROG	progressive
PRON	pronoun
PROP	proprietive
PROX	proximal
PRQ	prepositional quantifier
PRT	particle
PTCPL	participle
PUNCT	punctual
Q	question marker
R	realis
REC	recent
RECIP	reciprocal
RED	reduplication
REFL	reflexive
REL	relative
RELP	relative pronoun
REM	remote
RI	ruptured intent
RLT	relational particle
RP	realis past-present (nonfuture)
RPT	repetition
S	subject
SEND	send versatile
SEQ	sequential
SG	singular
SGOBL	singular oblique bound pronoun
SIM	simultaneous
SM	subject marker
SML	semelfactive
SPE	specific
SRF	subject referent in focus
SS	same subject marker
STAT	stative
SUB	subessive
SUBJ	subjunctive
SUFF	suffix
SUP	superablative
sv	*sub vocem* (for dictionary entries)
TOD.PAST	today's past tense
TOP	topic
TOW	towards subject
TRI	trial
UP	verbal extension indicating movement upward
V	verb
VEN	venitive
VIC	verbal inflection clitic
VIS	visual
VZR	verbalizer
W	witnessed

Abbreviations for sign language names

ABSL	Al-Sayyid Bedouin Sign Language
AdaSL	Adamorobe Sign Language
ASL	American Sign Language
Auslan	Australian Sign Language
BKSL	Ban Khor Sign Language
BSL	British Sign Language
CSL	Chinese Sign Language
DGS	German Sign Language (*Deutsche Gebärdensprache*)
ENG	Greek Sign Language (*Elliniki Noimatiki Glossa*)
GSL	Ghanaian Sign Language
HKSL	Hong Kong Sign Language
HPSL	Hawai'i Pidgin Sign Language
HSL	Hausa Sign Language
IPSL	Indo-Pakistani Sign Language
ISL	Israeli Sign Language
KK	Kata Kolok
KSL	Kenyan Sign Language
LIB	Lebanese Sign Language (*Lughat al-Isharat al-Lubnaniya*)
LIS	Italian Sign Language (*Lingua dei Segni Italiana*)
LIU	Jordanian Sign Language (*Lughat il-Ishaarah il-Urduniah*)
LSA	Argentine Sign Language (*Lengua de Señas Argentina*)
LSB	Brazilian Sign Language (*Língua de Sinais Brasileira*)
LSC	Catalan Sign Language (*Llengua de Signes Catalana*)
LSE	Spanish Sign Language (*Lengua de Señas Española*)
LSF	French Sign Language (*Langue des Signes Française*)
LSM	Mexican Sign Language (*Lengua de Señas Mexicana*)
LSQ	Quebec Sign Language (*Langue des Signes Québécoise*)
MSL	Maritime Sign Language
NGT	Dutch Sign Language (*Nederlandse Gebarentaal*)
NS	Japanese Sign Language (*Nihon Shuwa*)
NZSL	New Zealand Sign Language
OBSL	Old Bangkok Sign Language
OCSL	Old Chiangmai Sign Language
OFSL	Old French Sign Language
ÖGS	Austrian Sign Language (*Östereichische Gebärdensprache*)
PISL	Plains-Indian Sign Language
PSL	Providencia Sign Language
RSL	Rennellese Sign Language
RZY	Russian Sign Language (*Russkii Zhestovyi Yazyk*)

SKSL	South Korean Sign Language
ST	Swedish Sign Language (*Svenska Tecknenspråket*)
SV	Finnish Sign Language (*Suomalainen Viittomakieli*)
TCE	Irish Sign Language (*Teanga Chomharthaíochta na hÉireann*)
TİD	Turkish Sign Language (*Türk İsaret Dili*)
TSL	Thai Sign Language
TZS	Taiwanese Sign Language (*Taiwan Ziran Shouyu*)
UKSL	Urubú Sign Language (*Urubú-Kaapor Sign Language*)
USL	Ugandan Sign Language
VGT	Flemish Sign Language (*Vlaamse Gebarentaal*)

Database acronyms

APiCS	The Atlas of Pidgin and Creole Language Structures (http://lingweb.eva.mpg.de/apics/index.php/The_Atlas_of_Pidgin_and_Creole_Language_Structures_%28APiCS%29)
ASJP	The Automated Similarity Judgement Program (http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm)
LCGB	Language Convergence and Grammatical Borrowing Database
UPSID	UCLA Phonological Segment Inventory Database (http://web.phonetik.uni-frankfurt.de/upsid.html)
WALS	The World Atlas of Language Structures (http://wals.info/)

Languages cited in Chapter 1

1 Hawai'i Creole English
Hawaiian
2 English, American
3 Portuguese, Brazilian
4 Portuguese
5 Faroese
6 British Sign Language
English
7 French
8 Dutch
9 Swedish
10 Greek
11 Russian
12 Persian

Chapter 1

Introduction

Language is one of the most fundamental defining features of the human being, and scholars have been trying to capture its essentials for thousands of years. One way of doing this is to try to establish the limits and possibilities of human language by comparing the structures of a wide range of diverse languages. That is what linguistic typology is about and what this book is about. This introductory chapter will give the briefest of overviews of some key milestones in the history of linguistic typology (1.1) before bringing up the purpose of this book (1.2). In 1.3 I explain the general conventions I am using and in 1.4 I outline the structure of this book.

1.1 Fast forward from the past to the present

Linguistic typology is the systematic study and comparison of language structures, a practice that in essence goes back at least two and half centuries. As early as 1772, Johann Gottfried Herder discusses the merits of comparing languages in order to understand the speakers and their mentalities in his *Abhandlung über den Ursprung der Sprache* (*A Treatise on the Origin of Language*). One of the first linguists to propose a typological framework for characterizing language types was Friedrich von Schlegel, who contended that languages will have different strategies for organizing the linguistic devices they employ when associating various concepts with each other (von Schlegel 1808). This essentially led to a classification of languages according to a system of morphological types, which has been continuously revised since then.

The first to coin the term *typology* as a branch of linguistics was von der Gabelentz (1901 [1891]), who argued that classification of languages on a genealogical basis was not to be equated with classification of languages based on linguistic types, and for the latter approach suggested the term 'typology': "Dürfte man ein ungeborenes Kind taufen, ich würde den Namen Typologie wählen" ("If one were permitted to christen an unborn child, I would choose the name Typology"; von der Gabelentz 1901 [1891]: 481). Furthermore, he argued that there was no scale according to which languages could be more perfect than others, innovatively stepping away from the evaluative notions that had previously been associated with language types.

Edward Sapir essentially continued and further developed von der Gabelentz's approach. In his highly influential *Language: An Introduction to the Study of Speech*

(1921) he "rejects any kind of evaluative typology" (Graffi 2010: 35) and also argues that the classification of languages into morphological types is too simplistic, since one language can make use of several different strategies. Instead, he argues that classification must be based on "the nature of the concepts expressed by the language" (Sapir 1921: 136). These conceptual types may be expressed by different morphological techniques and to different degrees of synthesis. Sapir thus established that, since typology is a combination of features, no one language can be classified as belonging to a specific linguistic 'type'.

Linguistic typology took off in its modern form with the ground-breaking research of Joseph Greenberg, such as, for example, his seminal paper on a cross-linguistic survey of word order leading to a series of implicational universals (Greenberg 1963). He was thus in essence inspired by and continuing the comparative tradition of Roman Jakobson (cf., for example, 1929 and 1958) and the Prague School, a collection of linguists who in the first half of the 20th century gave emphasis to analysing languages as systems of functional units. Greenberg also attempted to establish methods for quantifying typological studies, in order that linguistic typology could meet scientific standards (cf. Greenberg 1960 [1954]). Furthermore, Greenberg re-introduced the importance of studying the ways languages change, but with the emphasis that language change gives us possible explanations for language universals (cf., for example, Greenberg 1978).

Since Greenberg's pioneering efforts linguistic typology has grown exponentially and is, as any science, continuously being enhanced and redefined as to methods and approaches. The last few decades have seen the compilation of large-scale databases with the help of ever more refined technology, which have led to new insights as well as given rise to new methodological issues. As is aptly stated by Johanna Nichols, linguistic typology

> is on a roll at the moment and is likely to continue. By now, descriptive coverage of languages worldwide, computational tools and expertise, genealogical classification, and understanding of research design are adequate to support comparison not only on lookup characters, but on more complex and abstract characters. (Nichols 2007: 236)

As this book is meant for those who have a basic linguistic background but are new to linguistic typology, the epistemological routes the discipline has taken, however fascinating, may be difficult to follow before having any real idea of what the discipline is about in its current state. I thus refrain from delving further into a discussion about the history of linguistic typology. Accessible overviews of the history of linguistic typology are Ramat (2010) and Graffi (2010) with further references. The collection of chapters in Shibatani & Bynon (1995) provides a very thorough discussion on the history of and approaches to linguistic typology.

1.2 The purpose of this book

There are already several introductions to typology available. Comrie (1981b), which was subsequently revised and updated for the second edition (Comrie 1989), is a classic and in many ways set the tone for describing what the linguistic subdiscipline of typology is about and how it relates to other linguistic disciplines, notably Generative Grammar. Croft (1990) is another classic, which was extensively revised and updated for the second edition (2003), providing in-depth discussions on the unity and diversity of language, also, crucially, from a historical (or diachronic) point of view. Whaley (1997) is an exceptionally accessible introduction to the study of linguistic typology and universals and Song (2001) introduces not only core fields of typological studies but probably gives the most thorough discussion of all introductory works on the methodological issues and approaches related to typology as well as the applicability of linguistic typology outside the field itself.

However, all of these introductions appeared before the publication of the groundbreaking *World Atlas of Language Structures* (Haspelmath et al. 2005; henceforth WALS[1]) and the surge of large-scale databases and surveys that has followed in its wake. Furthermore, all of the above mentioned introductions concentrate mainly on morphological, syntactic and morphosyntactic features of language systems, leaving out such linguistic features as, for example, phonology and pragmatics. Moreover, as in much typological literature, sign languages are completely absent from the discussions of these features.

> Although it is increasingly being recognised that sign language research has much to offer for linguistic typology (cf. Dotter 2001), sign languages are still largely absent from typological studies, in particular in the context where they should most obviously be included, that is, large-scale typological surveys covering substantial samples of the world's languages.
>
> (Zeshan 2004a: 7)

This book is an attempt to remedy this state of affairs by way of including the findings in the WALS to the greatest possible extent, as well as the findings of other publicly available databases and by extending the discussion to all major descriptive levels of a linguistic system, ranging from contrastive segments (i.e. phonology) to pragmatic issues such as politeness, while also covering such topics as parts-of-speech as well as language change. In addition, sign languages have been systematically integrated throughout the book.

The purpose of this book is first and foremost to serve as a course book introducing the reader to the unity and diversity of human language. It is directed towards those readers who have a basic linguistic background, i.e. have at the minimum done

1. Unless otherwise specified, the acronym WALS refers to the 2011 edition of WALS Online.

introductory courses in general linguistics and phonetics/phonology, but have never heard of typology before. The book thus seeks to provide an overview of the various kinds of linguistic features that have hitherto received attention in the study of linguistic typology.

While a general linguistic background is presupposed, we all know that it can be both confusing and somewhat overwhelming to try to grasp a new topic. Therefore short definitions and explanations are given for even the most basic linguistic concepts when they appear in the discussion. That is, I go by the principle of explaining or defining the various notational and terminological tools I make use of, even if they are very fundamental. For the lucky few who feel neither confused nor overwhelmed, the repetition will not hurt, for, as my Latin teacher assured us, *repetitio est mater studiorum ('repetition is the mother of learning').*

It is important to keep in mind that this book in no way makes any claims to being exhaustive. Rather, it should be seen as a smorgasbord (more correctly: *smörgåsbord*) of the multifaceted nature of linguistic typology and the richness of linguistic systems as we know them, which will hopefully whet the appetite enough to spur the reader into further investigation. Every effort has therefore been made to provide a starting point for further study in the form of reference tips and examples for each issue or field discussed.

As mentioned above, I have attempted to integrate the findings in WALS to the greatest possible extent, as well as findings in other major databases. The reader is greatly encouraged to consult these databases as much as possible. I have also attempted to cover all the major areas of a linguistic system, that is, phonology, morphology, parts-of-speech, grammatical relations, syntax, speech acts and politeness as well as the fundamentally important topic of language change. The general structure of the book thus to a large extent follows the broad outline of a descriptive grammar. Each chapter contains a wide range of data, at least one feature map showing the global patterning of a selected feature as well as a map of the languages mentioned in the chapter, and suggestions for further reading. The further reading references are by force a selected few that are intended to serve as starting points for more information; priority has been given to more recent literature and literature that I think will be accessible for those who are not familiar with the topic at hand. It is hoped that this book will not only whet the appetite of the newcomer to typology, but also serve the linguistic community in general as well as the fieldworker in particular as a guide to the current state of knowledge of the linguistic unity and diversity in the world, since knowing what patterns we have discerned so far may serve as a guide for analysis and classifications of new or poorly investigated patterns. Furthermore, a general overview of various linguistic systems and how they pattern in the world may serve as a guide for historical linguists for plausible reconstructions of earlier stages of languages.

1.3 Conventions

I have included as many examples as possible as illustrations of the issues discussed. In doing this, every effort has been made to provide original examples from primary sources as an attempt to complement the more commonly known examples that tend to be used as illustrations for frequently discussed phenomena. Furthermore, any time a group of languages mentioned contains five languages or fewer, I have named the languages in question. This means that each chapter contains a considerable number of language names.

1.3.1 Some remarks on the languages cited in this book

In typology the **genealogical affiliation** of the language as well as the **area** it is spoken in are two very important factors to know about (see further Chapter 3). Therefore this information is provided the first time a language is mentioned in the book, according to the format 'language name (language family (language genus): location)'.[2] Thus English would be presented as follows: 'English (Indo-European (Germanic): UK)', meaning that *English* is the language name, that it belongs to the *Indo-European* language family, and further to the *Germanic* branch (genus) of that family, and that it is spoken in the United Kingdom. The classifications of languages are based primarily on the WALS Online 2011 edition and, for those languages not included in WALS, on the 16th edition of the *Ethnologue* (Lewis 2009). For more on classifications of languages and definitions of family and genus, see 2.2.

I have for the most part based the **language locations** on WALS and the *Ethnologue*. It should be noted that the WALS language locations are based on the situation of the world prior to the European colonial expansion. This means that English is located in the United Kingdom, Dutch (Indo-European (Germanic)) in the Netherlands, French (Indo-European (Romance)) in France, Portuguese (Indo-European (Romance)) in Portugal, and so on, despite the fact that these languages are now spoken in a much wider area and in many different locations. However, on occasion new varieties that

---

2. This is done consistently except in those cases where all the necessary information is stated in the text when the language is first mentioned. For example, in a sentence like "Adamorobe Sign Language in Ghana" the macro-data '(Sign Language: Ghana)' would be redundant. Similarly, in a sentence like "the Niger-Congo languages Balanta and Bambara" only the missing information, in this case the genera and locations, will be given. Likewise, the macro-data will be missing for the first instance a language is mentioned in the text if an example immediately follows, containing that same information. In the interest of reading flow I allowed myself one exception above with respect to Latin, which is an Indo-European Italic language that was spoken in present-day Italy and beyond.

have emerged as a consequence of the European expansion diverge from the language spoken in the original, pre-colonial location. For example, Brazilian Portuguese may differ in certain respects from European Portuguese. If specific varieties are mentioned, the location given is an approximation of where that specific variety is spoken today. In order to provide an at-a-glance indication of where a language is spoken the location stated in the text is usually a country name, even though political borders do not necessarily mark linguistic borders.[3] The actual coordinates, however, tend to be of a major city in the region where the language is spoken, or, if the language is a national language, the capital city of the country. Thus the coordinates for English are London, the coordinates for Swedish (Indo-European (Germanic): Sweden) are Stockholm and the coordinates for Russian (Indo-European (Slavic): Russia) are Moscow, even though these capital cities are not in the centre of the regions where the languages are spoken. For languages where I base the location on information given in the *Ethnologue* or various descriptions, and if the region in question does not have any major city, the location is an approximation based on the information available. Locations for extinct languages are, for the sake of simplicity, given as present-day countries, even though that is not only an approximation, but also an anachronism.

Notice that **language maps** will contain only one dot per language, irrespective of how many speakers (if any) it has, and irrespective of how large the area it is spoken in is. This is because of two things: as a system, the language will provide information of human linguistic capacities irrespective of the number of speakers it has. It is essential to keep in mind that typology is about *linguistic systems* found in the languages if the world. The number of speakers of a language has more to do with historico-political events, such as expansion and political power due to technological strengths, than anything else. If we were to go by number of speakers, extinct languages, such as Latin, would never appear on any survey maps. The second reason for mapping languages with dots is that it would be impossible to make any visual sense of maps with multiple overlapping areas, since very few languages are spoken in well-defined, discrete areas. This of course means that the multifaceted nature that any language will have, comprising regional and social variation, gets reduced to one single dot. I cannot stress strongly

3. While I give only one country name in the text in order to not disturb the reading flow too much, the location given in the Appendix 2 language list will list all the major locations, i.e. may list several countries for one language. On a few occasions it makes very little sense to list the political country of a location for a language, such as Denmark for Faroese (Indo-European (Germanic)) or USA for Hawaiian (Austronesian (Oceanic)), even though that is, strictly speaking, correct. Most Faroese speakers, however, live on the Faroe Islands (which belong to Denmark), and most Hawaiian speakers live on the islands of Hawai'i (which belong to the USA). In such cases I have opted to give the intuitively more meaningful location in the text.

enough that this is a necessary compromise for large-scale surveys and does not in any way imply that a typologist views languages as monolithic systems.

Languages vary radically with respect to **number of speakers**, and quite a few of the languages cited in this book are nearly or perhaps already extinct. A collected list of all the languages mentioned in the book and some macro-data for them can be found in Appendix 2, including number of speakers as stated in the 16th edition of the *Ethnologue*. It should be kept in mind that the information stated in the *Ethnologue* may be of varying age, which means that for some languages the information is by now quite outdated. For example, a language that had only a handful of speakers in 1983 may well have become extinct by now, nearly 30 years later. Notice also that, unless otherwise specified, the population figures given are for the entire number of speakers listed, spanning all countries where the language is spoken. That means, for example, that the population figure given for English vastly exceeds the number of inhabitants in the United Kingdom, since English is spoken in many other places than there.

The **language names** given in this book are primarily based on those given in WALS, whose editors have attempted to give the names by which the languages are currently known, since older language names can be considered offensive by the speakers of the language. This means that some of the language names may occasionally differ from those used in other sources. For example, Romani dialects and speakers were once known as 'Gypsies', a denomination now considered quite offensive and, if anything, are only used pejoratively or as an insult and which should be avoided. In case a language name contains a modifier, to indicate, for example, that it is a specific variety of a language, or that it is a specific language of a group of languages, I give the name with the modifier initially in the text, but list it according to the larger language or group in Appendix 2. Thus, for example, *Brazilian Portuguese* and *American English* will be stated as such in the text, but listed as *Portuguese, Brazilian* and *English, American* in the appendix. Observe that not all languages containing several words are names plus modifiers. *Hawai'i Creole English*, for example, is the full name of the language, as is, for instance, *British Sign Language*, which means they will be listed as such in the appendix.

Sign language names are most commonly cited with abbreviations in the literature, often with the abbreviation of the native language name of the sign language, even if the text is written in English, such as DGS for *Deutsche Gebärdensprache* (German Sign Language). I have followed this convention. In order to familiarize the reader with sign language abbreviations, I have consistently given them the first time a sign language name appears, even in those cases where the name appears only once in the text.

1.3.2 Some remarks on the examples in this book

Most examples provided in this book are from languages that only the exceptionally few will have first-hand knowledge of. Therefore all examples, except the English ones, are consistently provided with morpheme-by-morpheme analyses and glosses, a practice called **interlinearization**, **interlinearized morpheme translation** or **interlinearized glossing**, among other terms. What this essentially means is that each relevant segment of the example is provided with a gloss. A **gloss** is basically an analytical explanation of the unit in question. For example, the English sentence *The dogs are chasing the cats* could be interlinearized as follows:

English (Indo-European (Germanic): UK)

(1)	the	dog-s	are	chasing	the	cat-s
	DEF.ART	dog-PL	be.PRES.PL	chase.PROG	DEF.ART	cat-PL

In (1) each word is provided with at least one gloss. The grammatical information is given as abbreviations (for a full list of the abbreviations used in this book and their full forms, see List of Abbreviations) and in small caps, while the bare form of the lexical information is given in full and in lower case. Thus the first word in the sentence is glossed as a definite article (DEFinite.ARTicle), the second word in the sentence is segmented into its lexical stem *dog* and the grammatical suffix *-s* indicating plural number, which is shown in the glossing by separating the two units with a hyphen. If a unit is not easily segmentable, the lexical information is given first, with the grammatical information following. Each component of the non-segmentable analysis is separated with a full stop; cf. *are* and *chasing* in the example above.

The glosses generally follow those set up in the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php). For more information about the principles and premises of interlinear glossing, see Lehmann (2000). Different authors may use different terms and abbreviations; while I have in almost all cases followed the analyses of the sources, I have streamlined the terms and glosses to be consistent throughout this book, making the examples more cross-compatible with each other.

It is very important to keep in mind that the **translations** given for each example are only the closest idiomatic equivalences to English. That is, a translation should never be seen as an analysis of data, but merely the closest approximation of language A into English. Since languages differ radically with respect to their characteristics, various kinds of information are often lost in translations. Drawing conclusions about the structure of a language based on translations would therefore be a highly questionable endeavour. For example, German has four different cases, nominative, accusative, dative and genitive, as well as three different genders, masculine, feminine and neuter.

This is not so in English and would therefore not show in the translations of German into English, although it would show in the glossing. Consider the following:

German (Indo-European (Germanic): Germany)

(2)	der Mann	the man	die Frau	the woman	das Kind	the child
	den Mann	the man	die Frau	the woman	das Kind	the child
	dem Mann	the man	der Frau	the woman	dem Kind	the child
	des Mann-es	the man-s	der Frau	the womans	des Kind-es	the child-s

If we base our analysis only on the English translation of the data above, we would come to the conclusion that German, just like English, has morphological marking for only one case, namely the genitive, which in German can even occasionally be left out (cf. the last row in the third column above). We would also come to the conclusion that there is in German, just like in English, no nominal gender. Both of these conclusions would be entirely wrong, as can already be surmised from the differences in form between the German definite articles. A glossing would show that the first row of each column above is in the nominative case, the second in the accusative, the third in the dative and the fourth in the genitive. Furthermore, glossing of the data would show that each column represents a different nominal gender, the first one masculine, the third one feminine and the fifth one neuter. In other words, it is important to keep in mind that the glossing provides the analysis of the data, while the translation is simply an idiomatic English approximation that serves to give an idea what the example means. It is also important to keep in mind that the glossing is not the data itself, but an analysis of it. Thus different researchers might analyse the same data differently and consequently gloss the same data differently. Furthermore, the same data may be glossed with varying degrees of detail depending on what the author is focussing on or considers relevant for the discussion.

In order to make examples immediately accessible, each is given with the language name and macro-data (i.e. affiliation and location) irrespective of whether this macro-data has already been provided in the text. For each example the source of the example is given, allowing the reader to make further inquiries about the language or example in question. Where I have not been able to go back to the primary source myself, I have still included a reference to it in order to allow the reader to trace the data. In those cases where I have based the example on my own knowledge or my own fieldwork of the language, this is indicated by the source reference 'personal knowledge' or 'own fieldwork' respectively.

Examples of sign languages are interlinearized according to more or less the same principles as the examples of spoken languages. By general convention glosses for the signs are in capital letters, not in lower case. I have followed this convention throughout the book. Thus the sign meaning 'apple' is glossed APPLE. Sign languages provide

a host of information with other means than the hands, so-called non-manual signs. These are usually indicated with a line above the signs that they accompany, with the grammatical analysis indicated with lower case abbreviations. A non-manual marker for negation accompanying the first half of a complex sentence, for example, would be indicated as follows:

Constructed example:

(3) _____ neg
APPLE EAT ALLERGIC
'He doesn't eat apples, (because) he's allergic.'

In (2) the glossing indicates that the signs made are 'apple', 'eat' and 'allergic' respectively, and that the non-manual marker for negation accompanies the two first signs.

1.4 The structure of this book

As mentioned above, the general structure of this book follows that of a descriptive grammar, starting with the smallest linguistic unit and moving ahead to larger linguistic units. The chapters in general build on each other, which means that I have to the greatest possible extent tried to avoid using terminology and concepts not previously introduced and explained. However, occasionally discussions have to be made without providing background explanations for terms, most notably in the introductory chapters.

There is no set model for which headings various linguistic features should be discussed under, and different sources will organize their discussion differently. The motivation for my organization is first of all to introduce issues gradually and in such a way that the discussion can lean on previous chapters and sections. Secondly, as the purpose of this book is to function as a course book, I have tried to make the chapters of roughly comparable lengths, although different topics by necessity demand different treatments and amounts of space. The chapters have a largely uniform format, starting with a short overview of what the chapter is about, then dealing with the topic at hand (including the pertinent issues relating to sign languages) and concluding with a summary. For each chapter a map of the languages cited in that particular chapter has been provided, as well as at least one feature map graphically showing the global patterning of a selected linguistic feature. The feature maps were generated with the 'Interactive Reference Tool (WALS program)' developed by Hans-Jörg Bibiko, provided with the printed version of WALS (Haspelmath et al. 2005) and available for free at http://www.eva.mpg.de/lingua/research/tool.php. The tool includes a guide on how to use it as well as such information as, for example, how to generate one's own maps.

Throughout the book I have included what I think of as 'curiosity boxes', that is, little insets bringing up rare phenomena or otherwise eye-catching issues not discussed in the text. These boxes are meant as glimpses of the fascinating mosaic that human languages have to offer. In order to make these 'curiosity boxes' immediately accessible, I have given the affiliation and location of each language mentioned in the box, irrespective of whether this macro-data has already been provided in the text.

Chapter 2 gives a definition of linguistic typology and language universals. It discusses the issue of classifications and why so-called 'contact languages' (pidgins, creoles and mixed languages) as well as sign languages tend to be treated separately. **Chapter 3** brings up the methodological issues of data, language samples and databases, as well as the issue of language endangerment, language documentation and description, and methodological issues in sign language typology. These two chapters provide necessary background information for those sections dealing with actual data, which forms the bulk of the book.

The remainder of the book deals with cross-linguistic patterns of language data. **Chapter 4** gives an introduction to the basic concepts needed for studying contrastive segments as well as syllables and suprasegments before moving on to describe various phonological patterns found across the world, including sign languages. Phonology has not previously been included in introductions to typology. **Chapter 5** brings up the notions needed to understand discussions on morphology, or the building blocks of language, before giving an overview of the different types of morphological strategies known in the languages of the world, including sign languages.

In **Chapter 6** I first discuss by what means the lexicon tends to be enriched, that is, various kinds of word-formation. I then discuss the notion of parts-of-speech, i.e. word or lexical classes, and which lexical classes we tend to find in the languages of the world, both spoken and signed. With the exception of Whaley (1997), systematic discussions on parts-of-speech have not previously been provided in introductions to typology.

Chapter 7 deals specifically with various processes involving noun phrases in spoken and signed languages. After having defined what a noun phrase is, I discuss the grammatical categories of number and noun class (or gender) before moving on to discuss syntactic processes involving the noun phrase. **Chapter 8** then moves on to define the verb phrase and specifically characterize and discuss the grammatical categories of tense, aspect, and mood and modality in spoken and signed languages.

In **Chapter 9** I give an overview of the core units and processes involved in simple clauses, also in sign languages. This leads on to **Chapter 10** where a brief introduction to syntax (i.e. the building blocks of sentences) is given before discussing the way languages of the world, including sign languages, tend to organize their constituents, both on a clausal and a phrasal level.

In **Chapter 11** I discuss various kinds of complex clauses, bringing up the notions of and exemplifying coordination, subordination and cosubordination in languages of the world, including sign languages.

Chapter 12 gives an overview of various pragmatics topics in the linguistic systems across the world for both spoken and signed languages, and brings up not only different kinds of speech acts, but also the notion of politeness and how that may affect the structure of a language. Pragmatics has not received much attention in introductions to linguistic typology; this is especially true for the linguistic domain of politeness.

If there is one indisputable absolute linguistic universal, it is that all languages change. **Chapter 13** therefore brings up the topic of internal versus external change in both spoken and signed languages, specifically grammaticalization or how grammatical categories enter the linguistic system, and contact-induced change as well as linguistic areas. With the exception of Croft's (2003) discussion on diachronic typology, these two major domains relating to the evolution of linguistic structures have largely been left out in previous introductions to typology.

Appendix 1 provides a selected sample of sites containing much information and material of interest to anyone dealing with linguistic typology, both online databases and other kinds of sites. This is by no means an exhaustive list of sites, mainly a starting point for the interested reader to investigate and move on from.

Appendix 2 is a list of all the languages mentioned in the book, with some macrodata given for each language, such as genealogical affiliation, approximate location, number of speakers and the language code in the 16th edition of the *Ethnologue* (Lewis 2009), which will allow the interested reader to investigate further.

Finally, a **glossary** gives short definitions of the terms that appear in boldface in the book.

1.5 Keywords

gloss
interlinearization
language locations
language names
linguistic typology
translation

1.6 Exercises

1. Which information is lost in the translation from Greek (Indo-European (Greek): Greece) below?

enan	pássalo	~	mia	trapeza
ART.M.SG.ACC	pole.M.SG.ACC		ART.F.SG.NOM	bank.F.SG.NOM
'a pole'			'a bank'	(adapted from Ruge 1984)

2. Using the List of Abbreviations, how would you gloss the sentence below?

   The girls were eating sausages.

3. Identify the non-manual marker in the example below and show which part of the sentence it refers to.

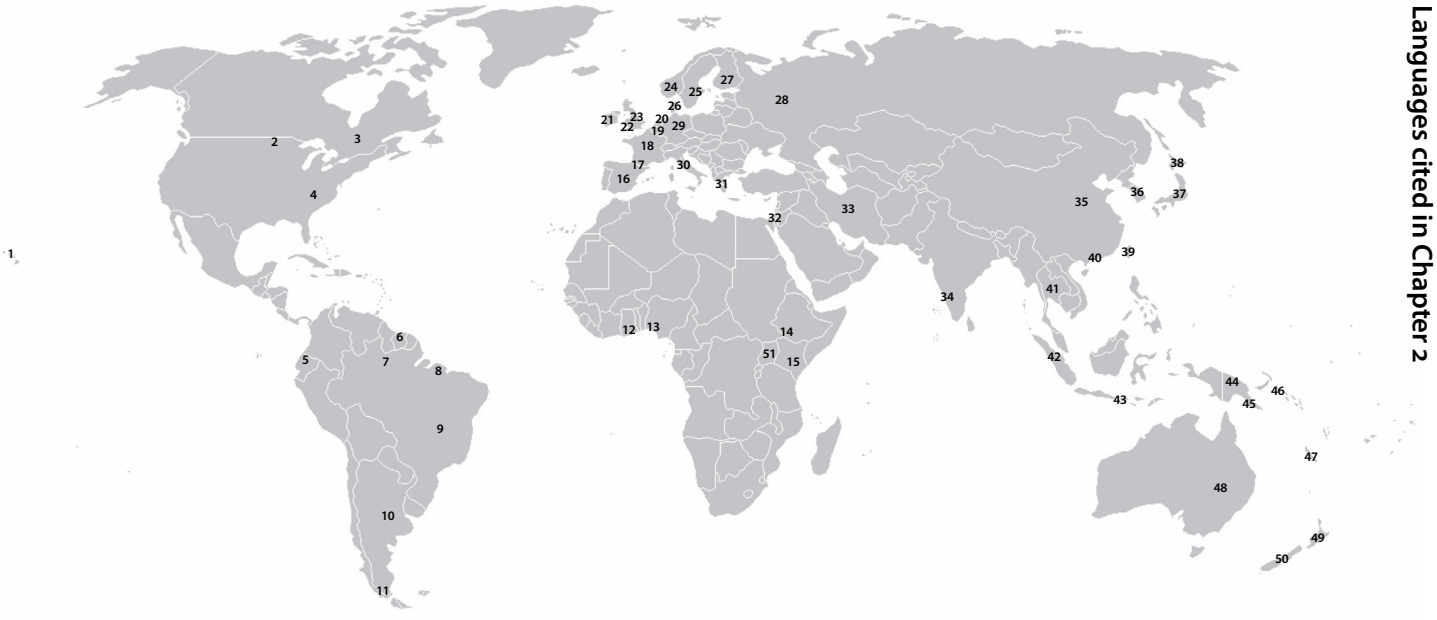
   _____neg
   LOOK CAT SEE
   'I looked, (but) I didn't see (any) cat.'

4. Why don't languages with a higher number of speakers get bigger dots on the language maps?

5. Is the following statement true or false? Motivate your answer.

   Typologists assume that languages are monolithic systems with no regional or social variation.

Languages cited in Chapter 2



1 Hawai'i Creole English
2 Michif
3 Quebec Sign Language
4 American Sign Language
5 Media Lengua
Quechua, Imbabura
6 Sranan
7 Hixkaryana
8 Urubú Sign Language
9 Brazilian Sign Language
10 Argentine Sign Language
11 Selknam
12 Adamorobe Sign Language
Ghanaian Pidgin English
13 Nigerian Pidgin English
14 Aari
15 Kenyan Sign Language
16 Spanish
17 Basque
18 French Sign Language
19 Flemish Sign Language
20 Dutch Sign Language
21 Irish Sign Language
22 Welsh
23 British Sign Language
English
24 Norwegian
25 Swedish
Swedish Sign Language
26 Danish
27 Finnish Sign Language
28 Russian Sign Language
29 German
German Sign Language
30 Italian
Italian Sign Language
31 Greek
32 Hebrew
Israeli Sign Language
33 Persian
34 Konkani
35 Chinese Sign Language
Mandarin
36 Korean
South Korean Sign Language
37 Japanese
Japanese Sign Language
38 Ainu
39 Taiwanese Sign Language
40 Hong Kong Sign Language
41 Thai Sign Language
42 Malay
43 Kata Kolok
44 Yimas-Arafundi Pidgin
45 Tok Pisin
46 Rotokas
47 Tayo
48 Australian Sign Language
49 Maori
50 New Zealand Sign Language
51 Ugandan Sign Language

Chapter 2

Typology and universals

Typology is the study of linguistic systems and recurring patterns of linguistic systems. Universals are typological generalizations based on these recurring patterns. This chapter will give a definition of what typology is (2.1). Section 2.2 brings up the notion of classifications, especially genealogical, and gives a brief discussion on why pidgins, creoles and mixed languages, as well as sign languages tend to be treated separately (2.2.1 and 2.2.2 respectively). In 2.3 I give a definition of universals and bring up the concepts of unrestricted (2.3.1) and implicational (2.3.2) universals. Section 2.4 takes a cursory glance at the motivations for language universals.

2.1 What is typology?

To put it very simply, linguistic typology concerns itself with the study of structural differences and similarities between languages. The term **typology** is, as many other linguistic terms, borrowed from the field of biology and means something like 'taxonomy' or 'classification' (Croft 2003: 1), or, to be precise, "the study and interpretation of types" (Pearsall & Trumble 1996: *sv*). Linguistic typology, then, is the study and interpretation of linguistic or language types. More specifically, it is the study and interpretation of types of linguistic systems. While this may involve comparison of linguistic systems within a language, it more generally involves comparison of linguistic systems between languages. Linguistic typology can be both **synchronic**, i.e. a comparison of languages contemporary to each other, or **diachronic**, i.e. a comparison of languages at various stages of their historical development. Impressionistically, synchronic typology has received more attention than diachronic typology, but both are equally necessary and can be thought of as complementary to each other (cf. the discussion in Croft 2003: 232ff).

Any linguistic system may serve as a starting point for typological comparisons. Thus we may, for example, have typological surveys of phonological, morphological, grammatical, syntactic, lexical, pragmatic, semantic, etc. systems. Essentially, the ultimate goal of linguistic typology "is to understand [the question] 'what's where why?'" (Bickel 2007: 248). In other words, a driving force is to try to establish recurring patterns across languages, in order to answer the questions "what is out there?", "where does it occur?" and "why do we have particular patterns?". If we want to formulate

hypotheses about the unity, diversity, potentials and limits of human language, we need to know what human language is capable of. Investigating only one language will not be sufficient to answer such questions. For instance, if we look at English we may establish various linguistic factors, such as what the phoneme inventory is, what the morphology is like, what grammatical categories we can discern, how units are ordered, and so on. We may then use these factors to hypothesize what the human brain needs or does not need in order to allow a person to produce and maintain a coherent language and communicate with others. Based on English, this might lead us to assume, for example, that a human language needs quite a number of vowels in order to get by. We might further assume that the only way to know what grammatical functions the words in a sentence have depend on how they are ordered with respect to each other. Compare, for example, the following:

(4) a. John called Mary.
b. Mary called John.

In (4a) *John* is the person who did something (*John* is the subject of the clause), and we know that because *John* is placed before the verb *called*. *Mary* is the object of the clause because *Mary* is placed after the verb. If we swap *John* and *Mary* the grammatical relations also swap and *Mary* becomes the subject while *John* is the object, as in (4b). Based on English, we would thus assume that word order follows a rigid pattern of subject-verb-object. We might also assume that a subject must be expressed in a clause for the clause to be grammatically acceptable, even if there is no physical entity to be referred to. In a sentence like *He swam*, for example, *He* refers to some (male) human or animal. *He* is the subject of the verb *swam*. But in a sentence like *It rained*, we have a neutral pronoun filling the slot of the subject, even though *It* does not refer to anything. Taking out the subject, however, is not possible; a clause like **Rained* is not grammatically acceptable.[4] This might lead us to conclude that the human brain demands that every clause has at least one slot for "subject" and one for "verb" in order to be complete, even if the "subject" slot is filled with a semantically empty reference (i.e. even if the reference does not have any concrete real world meaning).

Based on some other language, we might get an entirely different picture, which would lead us to make an entirely different set of assumptions. If we base our hypothesis on what a human language must have on Mandarin, for example, we would assume that we only need a handful of vowels to get by, but that differences in tone is an absolutely essential requirement of human language. Compare the four words in (5):

4. The symbol * means that the form or construction is non-existent or grammatically unacceptable.

Mandarin (Sino-Tibetan (Chinese): China)

(5) a. yī 'cloth' b. yí 'to suspect'
c. yǐ 'chair' d. yì 'meaning' (Li & Thompson 1990: 814)

As (5) shows, the choice of tone (indicated on the vowel by the diacritics ¯, ´, ˇ and ` respectively) is an essential part of conveying the meaning of the word in Mandarin. Furthermore, if we base our hypothesis on Mandarin only, we might assume that copying parts of the word, a process called reduplication, is an essential requirement for a language to be able to form adverbs from adjectives:

Mandarin (Sino-Tibetan (Chinese): China)

(6) màn 'slow' → màn-màn-de 'slowly' (Li & Thompson 1990: 821)

Both tone and reduplication are largely irrelevant for the linguistic system of English.

Based on yet another language we would make yet different assumptions. In Modern Greek, for example, the order of the elements in the clause is not rigid. Compare the sentences in (7) below:

Modern Greek (Indo-European (Greek): Greece)

(7) a. o skilos kiniyinuse ti yata
the dog chased the cat

b. kiniyinuse o skilos ti yata
chased the dog the cat

c. o skilos ti yata kiniyinuse
the dog the cat chased

d. ti yata o skilos kiniyinuse
the cat the dog chased

e. ti yata kiniyinuse o skilos
the cat chased the dog

f. kiniyinuse ti yata o skilos
chased the cat the dog
'The dog chased the cat.' (Ruge 1984: 94)

In (7) the elements of the sentence (*the dog*, *the cat* and *chased* respectively) may move around in relation to each other. The role of subject always stays with *o skilos* 'the dog', because *o skilos* is morphologically marked for the nominative case. The role of object always stays with *ti yata* 'the cat' because *ti yata* is marked for the accusative case. Based on Greek we would therefore assume that while the order of the elements in the sentence is largely irrelevant, morphological devices to indicate the case of the nouns are essential tools for humans to be able to sort out what grammatical relations

the elements have. Case marking is largely irrelevant for English and Mandarin. Furthermore, based on Greek, we would conclude that an overt reference to a subject is not necessary if the subject is not a real world entity:

Modern Greek (Indo-European (Greek): Greece)

(8) θa vréksi
FUT rain.3SG
'(It) will rain.' (Joseph & Philippaki-Warburton 1987: 23)

In (8) there is no noun or pronoun referring to 'rain' or 'it' (*θa* is a marker for tense and belongs to the verb). In Greek, only an inflected form of the verb 'rain' is necessary for the sentence to be grammatically correct. This would not be possible in either English or Mandarin.

What I have tried to show here is that if we look only at one single language, or maybe a very small number of languages, and base our hypotheses on what humans require, or tend to need, or tend to avoid in their communicative tool called language, we are likely to end up with assumptions that would not hold, given that languages can differ from each other a great deal. Looking at a larger number of languages we would, for example, see that, contrary to English, languages can easily get by with very few vowels, or that some languages demand a very large inventory of consonants indeed to differentiate between different meanings. We would find that some languages employ very little or no inflectional morphology, while others demand extremely complex kinds of inflections for sentences to be grammatical. We would see that while some languages have a rigid word order, others do not. In other words, if we want to be able to answer the fundamental question "what is language?" then we will have to know what kinds of solutions speakers have found to encode meaning, i.e. what kinds of systems their languages have ended up acquiring. In order to get hold of such information, we need to compare between languages across both space and time.

Linguistic typology thus often involves **cross-linguistic comparison**, i.e. comparison between different languages.[5] While the term 'cross-linguistic' as such simply means "across languages" and can be used for a comparison between only two languages, I use it here and throughout this book to imply across several languages. With cross-linguistic comparison linguistic typology can give an idea of how linguistic features pattern across the world. Some solutions are common, some are rare. Typology

5. This is often contrasted with Generative Grammar, which approaches the study of language by looking very closely indeed at one language or a small number of languages. Different approaches to the study of language all have their merits and complement each other in the common quest to try to understand the fundamentals of human language. For an accessible introduction to Generative Grammar, see, for example, Carnie (2002).

can map the patterns, which may then serve as a starting point for investigations into why we find those kinds of patterns. Typology can also serve as a guide to analysis of languages.

> Once there is a clear and precise classification of occurring patterns, a new pattern may be evaluated with respect to existing ones. In diachronic [i.e. historical] analysis, where pieces of the puzzle (living speakers, phonetic studies of them, etc) may often be missing, typological work can be particularly useful in guiding the analysis in one direction over another. (Blevins 2007: 110)

While the quote above focuses specifically on the importance of typology for explaining sound patterns, the statement holds for any linguistic domain.

It is important to note here that typology is not able to establish what is *possible* in human language, for the simple reason that not all human languages are available for analysis. For one, it is not possible to include all of the roughly 7000 languages currently known to exist in the world (Lewis 2009), because we do not have adequate descriptions for all of them. But even if we did have descriptions for all the known languages in the world, we do not have access to all languages that have already gone extinct, nor do we have access to those languages that are yet to emerge. Thus typology is about probabilities more than anything else. In short, linguistic typology is basically about tendencies and by extension trying to explain why we get these tendencies.

2.2 Classifications

A central concept in typology is **classification**. Languages can be classified according to various criteria. We could choose to group languages by the number of speakers they have, or by formality of the situation where the languages are used, or by the area they are spoken in, or by the genetic affiliation they have, and so on. With linguistic typology the classification is primarily based on the elements that make up the structure of languages, such as sounds, words and parts of words, how these words are organized with respect to each other, and so on. Thus we could say that Mandarin is a reduplicating language (while English is not) and is also a tone language (while English is not). We can say that both English and Chinese are subject-verb-object languages. And so on. When investigating correlations between languages, we may, for example, investigate whether reduplicating languages also generally tend to be tone languages (they do not, in fact; most of the reduplicating languages in Rubino's (2011) database are listed as languages lacking tone in Maddieson's (2011f) database). Since the bulk of this book is about what types of linguistic systems and patterns we tend to find across languages, that kind of classification will not be discussed further in this section.

Languages may, as mentioned, be classified according to their size. The languages of the world vary radically with respect to how many speakers they have. The vast majority of the languages of the world (94% to be exact, cf. Lewis 2009) are spoken by a population of less than one million people. The eight languages with the largest population size make up for almost two fifths of the population of the world. Most of the known languages are spoken by fewer than 10,000 people (Lewis 2009). Classifying languages by size, one may, for example, investigate whether linguistic complexity and population size of the language have any kind of correlation. For discussions on population size and linguistic complexity, see, for example, Dahl (2004) and the chapters in Sampson et al. (2009).

Languages may also be classified according to social factors, such as the age, gender, or level of education of the speakers, or the formality of the situation the language is used in, and so on. We must always remember that language is used in communication between individuals. Thus the social context and a number of social factors such as the image of the speaker and the situation of the language use play a role in how language is employed. And language is not monolithic. The same language will be used differently by the same user in different contexts. Likewise, the same language will be used in slightly different ways by different users. For example, speakers with a higher level of education may use their language differently from speakers with little or no education. Speakers of one generation may use their language differently from speakers of another generation. Women may use their language differently from men. Or the formality of the situation may influence the language: it is unlikely that we speak in the same way when we give a presentation in front of an audience as when we are sitting at home in our pyjamas having breakfast (if we are at all inclined to speak in the latter situation). The way such social factors play a role in the structure of language is studied in sociolinguistics. For a very accessible introduction to sociolinguistics, see Meyerhoff (2006). For a discussion on how sociolinguistic factors correlate with linguistic typology, see Trudgill (2011).

There are many more ways in which we could group languages together. While the above kinds of classifications are relevant for typology, two types of classifications are especially essential for cross-linguistic surveys: the **genealogical affiliation** of a language and the **area** it is spoken in. This is because in order to make typological surveys as representative as possible of the world's languages, they need to be controlled for genealogical and areal bias, or the patterns that appear might be due to such factors as inheritance or language contact. For more on methodological issues, see Chapter 3.

Languages may be grouped together by the location where they are spoken. This is not as straightforward as it may seem. One question is, for example, whether to

group languages by countries they are spoken in. This may be useful for organizational purposes, but we must always remember that such things as country borders are recent politico-historical artefacts that do not necessarily represent historico-cultural or linguistic boundaries. It should therefore be kept in mind that a country name as a location for a language merely serves as a practical shortcut for giving a rough idea of where the language is spoken. Linguistic areas are more likely to develop where there is much contact between people, irrespective of political borders. When people are in contact, they are likely to affect each other's languages somehow (for more on language contact, see 13.2). People from areas that are remote and difficult to reach are less likely to have sustained contact with other people. Migration patterns, which would also be a factor of language contact, are likely to follow natural boundaries for the simple reason that one is likely to take the less arduous routes in search for new land. Very large areas with natural boundaries are continents, but within continents we may have natural boundaries such as inaccessible mountain ranges. One may, for example, investigate whether languages are likely to be more "unusual" or dissimilar from others if they are spoken in areas that are remote or difficult to access in various ways (in fact, Bickel 2006 has shown that languages in the mountainous areas of the Eurasian continent, namely the Caucasus and the Himalayas, tend to be typologically unusual). If so, it might be due to lack of contact with large migration waves in geographically more accessible areas that may have made languages affect each other and become more similar to each other.

Languages may also be grouped together according to their origins. Languages that descend from a common ancestor are grouped together into one language family. This is done by rigorous methods of historical and comparative linguistics. Historical (or diachronic) linguistics investigates how languages change over time. By comparing languages we may establish whether they are related or not, and if they are related, how close their relationship is. A common starting point is to look for potential cognates, i.e. words that share a common origin. Without going into details, by using the various techniques of the comparative method we proceed to group the languages according to how closely related they seem to be. This allows us, as a working tool, to construct family trees, where the parent language, from which the other languages originate, is at the top, with the descendants branching out according to our subgroupings. It also allows us to reconstruct plausible forms for earlier (usually unattested) stages in the family, so-called 'proto-forms'. In essence, with family trees we implicitly make the assumption that languages are split-off branches from earlier languages, i.e. that they are direct descendants from the earlier languages. Very simplified, we picture that a language is spoken, and then various factors such as internal change and distance from other speakers contribute to give rise to dialectal variation in the language. Eventually

the dialects become so different that they have developed into different languages.[6] In other words, family trees generally assume one parent language for several daughter languages. For a very accessible introduction to historical linguistics and the comparative method, see Campbell (2004).

Central concepts in terms of genetic affiliation for typology, as we shall see later (3.2), are the genus and the family of a language. The **family** of a language is the highest level of the affiliated languages. An example of a language family is Indo-European. Language families may be of radically different sizes and ages. The **genus** of a language is "a level of classification which is comparable across the world, so that a genus in one family is intended to be comparable in time depth to genera in other parts of the world" (Dryer 2011d). That is, all genera across the world are hypothesized to be of roughly the same age. The term, originally suggested by William Croft, mirrors the taxonomic level of genus in biology, which refers to an obviously closely related set of species (Dryer 1989a: 267). Examples of genera are Germanic, Celtic and Romance. Some languages cannot be demonstrated to have any genealogical relationship with any other language. These kinds of languages, called **language isolates**, can be thought of as belonging to a family which is made up of only one genus, which in turn consists of only one language. Examples of language isolates are Korean (North, South Korea) and Ainu (Japan). Some language isolates consist of varieties, or dialects, which are all mutually intelligible. An example of such an isolate is Basque (Spain).

Since there are hundreds of language families in the world, and even more genera that belong to these families, it would be impossible to list them all here. The ten major families are listed in Table 2.1 (ordered roughly by size):[7]

---

6. It is not possible to give an exact definition of what a language versus a dialect is. A common rule of thumb is that languages are not mutually intelligible, while dialects are. It is important to keep in mind that this is a matter of degree. Furthermore, extra-linguistic factors such as political borders also play a role. For example, Swedish, Danish and Norwegian are all mutually intelligible, but they are considered different languages and not 'merely' dialects. Italian dialects, on the other hand, are not necessarily mutually intelligible, but they are called dialects and not languages. It is thus best to view it as a continuum on a scale, where on the one end of the scale we have speakers using the same language (they understand each other without difficulties), on the other end of the scale we have speakers using different languages (they do not understand each other at all), and in the middle we have speakers using different dialects of a language (they understand each other to varying degrees). For more on languages versus dialects and the study of dialects in general (dialectology), see Chambers & Trudgill (1998).

7. Individual language families and how they are structured can be a controversial issue. I am basing my genealogical classification on Dryer & Haspelmath (2011) and Lewis (2009).

Table 2.1 The ten largest language families, ordered by size.

FAMILY	NO. OF LANGUAGES	APPROXIMATE AREA
Niger-Congo	over 1500	entire sub-Saharan Africa
Austronesian	over 1200	from Southeast Asia to Hawai'i; from Taiwan to New Zealand
Sino-Tibetan	ca 450	from the Himalayas to eastern China; from northern China to Thailand
Indo-European	close to 450	from Western Europe to Bangladesh; from Scandinavia to the Maldives
Trans-New Guinea	close to 400	Papua New Guinea and Indonesia
Afro-Asiatic	close to 400	Northern and Eastern Africa; the Middle East
Australian	over 250	Australia
Nilo-Saharan	ca 200	Central and Eastern Africa
Oto-Manguean	over 150	Mexico
Austro-Asiatic	over 150	from Eastern India to Vietnam; from Bhutan to Malaysia

It should be noted that languages that fall into the same structural type are not necessarily related or located in the same area. For example, both Welsh (Indo-European (Celtic): United Kingdom) and Maori (Austronesian (Oceanic): New Zealand) have the same basic word order of verb-subject-object (Dryer 2011r), but they are neither genetically related nor located anywhere near each other. Nor do the same origins or areal location imply that languages are necessarily of the same type. Welsh and English, for example, belong to the same language family and are located adjacent to each other, but have different basic word orders (the English basic word order being subject-verb-object; for more on word order typology, see Chapter 10).

2.2.1 A note on pidgins, creoles and mixed languages

Pidgins, creoles and other so-called '**contact languages**' form a group of languages that does not easily fit in any genealogical classification.[8] This is because they do not conform to the assumption that lies behind family trees, namely that a language is a (neat) split-off branch from an earlier language and therefore essentially only has one parent. Pidgins, creoles and other kinds of contact languages have more than

8. The term 'contact languages' is, in a sense, a rather meaningless term for two reasons: one, because no known language is immune to contact and contact-induced change (see further 13.2); and two, because languages such as English, Modern Hebrew and Persian, which have gone through periods of intense language contact, are not labelled 'contact languages'. In general the term is used to mean languages that have gone through intense language contact and that cannot easily be shown to descend from one single parent. It is in that sense the term is used here.

one parent. The cover term contact language is meant to capture the fact that pidgins, creoles and mixed languages emerged in situations of intense language contact. Very often the lexicon of a contact language mostly derives from one language, while the phonology, grammar and structure derive from other languages. Most often this occurred in situations where there was an urgent need to communicate and no common language to use. While the socio-historical situation that gave rise to an individual contact language is unique for each language, there are some common denominators. This section will give a very brief overview of pidgins, creoles and mixed languages. It is by necessity a highly simplified sketch of how these languages emerged and should only be taken as a very rough guide to the topic. For a very accessible introduction to pidgins and creoles, see Holm (2000) with further references. For background overviews on individual pidgins and creoles, see Holm (1989). A wealth of information on individual pidgin and creole languages can be found in Michaelis et al. (2013).

Urgent need for communication very often arises in situations of trade. The trading parties might not have any common language and therefore make use of some kind of a communicative bridge. If this communicative bridge is used more or less systematically, a common linguistic variety emerges. The intense and large-scale trading that the European exploitation and colonization of both the new and the old world involved gave rise to a number of contact languages. Not only goods were traded, but also people, which led to large-scale (forced or voluntary) migrations. Slaves or indentured labourers speaking different languages were brought to or sought jobs in plantations and in homes (to, for example, serve as domestic labour) and had to quickly be able to communicate. Very often the **target language**, i.e. the language that people aimed to learn, was the language of the colonizers. In other words, in English colonies, the target language tended to be English, while in Dutch, Spanish (Indo-European (Romance): Spain), French and Portuguese colonies the target language tended to be Dutch, Spanish, French and Portuguese respectively.[9]

There are no straightforward and uncontroversial definitions for 'pidgin', 'creole' and 'mixed language'. I stress once again that the sketch provided in this section is by necessity highly simplified.

A **pidgin** is a language that emerges when groups of people are in close contact and need to communicate but have no language in common. Typically the speakers with less power, for example slaves and workers, use the lexicon of those with more power, such as the colonizers. The language that provides the lexicon for a contact language

9. An exception to this was the situation on the plantations in the German colonies in the northern half of Papua New Guinea, Kaiser Wilhelmsland and Bismarck Archipel, where the German plantation owners initially communicated with their labour in Malay (Austronesian (Malayic): Malaysia) (because a large proportion of workers were recruited from Java), then later, when the labour became more mixed, in the pidgin English which became Tok Pisin (Pidgin (English-lexified): Papua New Guinea) (Huber 2009).

is called the **lexifier** language. The target language (or **superstrate** language) is usually the lexifier. The native languages that the speakers with less power have, the **substrate** languages, may influence the pidgin in various ways. All users, both those with more and those with less power typically cooperate

> to create a make-shift language to serve their needs, simplifying by dropping unnecessary complications such as inflections (e.g. *two knives* becomes *two knife*) and reducing the number of different words they use, but compensating by extending their meanings or using circumlocutions. (Holm 2000: 5)

A pidgin is thus typically a reduced language and nobody's mother tongue.[10]

A **creole** is typically a pidgin or jargon that has become the native language of a community. Very often this is the language of those whose parents or ancestors were displaced geographically for various reasons. With the ties to the original language and culture disrupted, the new generation that arises in these contact situations ended up nativizing the contact language and making it into their primary language. This process can be quite rapid; it is entirely possible that while the first generations of people born in the new location may have spoken one or both of their parents' native language(s) at home, they used the contact language when communicating with their peers.[11] In other words, playgrounds, schools and similar kinds of domains were probably instrumental in the formation of creoles, and this formation can be as rapid as a single generation. By making it into their primary language, these new generations also often extended the language. This process, called **creolization** (or **nativization**) is still a matter of much debate, and it is beyond the scope of this section to give proper attention to the various issues involved. The main point is that a creole is a full-fledged language that can fulfil any and all functions that a human language needs to fulfil. This is in contrast to a pidgin, which tends to be restricted in use to only the situation where it is specifically needed.

There are essentially two main ways that pidgins and creoles tend to be classified, by lexifier and by region. Grouping pidgins and creoles together by lexifier would yield such categories as, for example, English-lexified[12] contact languages, Portuguese-lexified

---

10. This is of course a simplification. Pidgins may also acquire native speakers, especially in urban environments, as is happening with Nigerian Pidgin English (Pidgin (English-lexified): Nigeria) and Tok Pisin (cf. Muysken & Smith 1995).

11. This is what happened in Hawai'i, as has been shown by Roberts (2000).

12. It is a matter of vivid discussion among creolists whether to analyse pidgins and creoles as dialectal varieties of their lexifier languages or whether to analyse them as discrete languages that have emerged in situations of extreme contact. It is beyond the scope of this book to discuss that issue. The use of the modifier -lexified (as in English/Dutch/French/Spanish/Portuguese-lexified) is here meant only to indicate what the main provider of the lexicon was, and makes no further claim as to the genesis of the languages.

contact languages, and so on. Based on lexifier, such languages as Ghanaian Pidgin English (Pidgin (English-lexified): Ghana), Sranan (Creole (English-lexified): Suriname) and Hawai'i Creole English (Creole (English-lexified): Hawaiian Islands) would be grouped together, even though the socio-historical backgrounds differ considerably between these languages. Alternatively, pidgins and creoles may be grouped together by region, such as Atlantic pidgins and creoles or Pacific pidgins and creoles. This would possibly mirror the socio-historical backgrounds better, since the way the populations moved or were displaced was largely dependent on how the seas were navigated for trade. For example, the regions on either side of the Atlantic Ocean were, among many other things, linked together by the transatlantic slave trade, while the regions of the Pacific were, among many other things, linked together by mass movements of workers signing up as indentured labourers on plantations or signing up on ships as workers in the whaling industry.[13] Based on region, such languages as Hawai'i Creole English, Tayo (Creole (French-lexified): New Caledonia) and Yimas-Arafundi Pidgin (Pidgin (Yimas-lexified): Papua New Guinea) would be grouped together, even though they each have different lexifiers. Most commonly pidgins and creoles are first categorized by their lexifiers, then by the region they are (or were) spoken in.

A different type of contact language is the so-called **mixed language**, which typically has two ancestors. The most common definition of mixed languages is that they "emerged in situations of community bilingualism" and thus have split ancestry (Matras & Bakker 2003: 1). This refers to different languages with quite varied socio-historical backgrounds. One thing that unites them is that their contact situation was usually different from that of pidgins and creoles. Very often the contact situation was one of rather stable bilingualism, either due "to mixed households accompanying the formation of new ethnic identities, or through rapid acculturation leading to the adoption of a hybrid group identity, or through continuous socio-ethnic separateness resisting pressure to assimilate" (Matras & Bakker 2003: 14). It is common for mixed language speakers to also have knowledge of the two input languages. In fact only one mixed language, Michif (French, Plains Cree: Canada),[14] is known to be used natively and "independently of speaker's knowledge of any of its source or ancestor languages" (Matras &

13. In fact, it might be interesting to investigate whether the different socio-historical and economical backgrounds of the Atlantic and Pacific trade regions entailed different kinds of demographic migrations. It could possibly have been the case that there was a more or less balanced ratio of men and women that got displaced in the slave trade of the Atlantic, while the indentured labour and especially the whaling industry might have led to a greater population movement of men in the Pacific. If that was the case, it would be interesting to investigate whether that might have made a difference in the outcome of the contact languages that emerged through these population movements.

14. Note that this classification follows that in APiCS and differs from that in WALS, where Michif is classified as an Algic Algonquian language.

Bakker 2003: 2). Thus, many speakers of Media Lengua (Ecuador), for example, which has its grammar from Imbabura Quechua (Quechuan (Quechuan): Ecuador) and its lexicon from Spanish, are trilingual and also know Quechua (used with Indians outside the community) and Spanish (used with non-Indians) (Muysken 2013).

In linguistics the name given to pidgin and creole languages may differ from the name the speakers themselves use for the language. In such situations the language name used by the linguist tends to reflect three things: the approximate location of the language, whether it is a pidgin or creole, and the lexifier. For example many English-lexified pidgin and creoles are called Pidgin by their speakers. Thus, in order to be able to distinguish between Pidgin in Ghana, which is an English-lexified pidgin spoken in Ghana, and Pidgin in Hawai'i, which is an English-lexified creole spoken in Hawai'i, the linguist may call the former Ghanaian Pidgin English and the latter Hawai'i Creole English. This can be a sensitive issue, as there has been a long history of condescension towards pidgin and creole languages, viewing them as "broken" or "incomplete" languages. It is very important to keep in mind that these are not inferior languages in any way. They are also not "bad" or "lazy" versions of other languages. They are simply languages in their own right, just as Japanese (Japanese (Japanese): Japan) or German are languages in their own right. It is also important to keep in mind that the language name used by the linguist is simply a shorthand version meant to clarify what language is being discussed and does not necessarily imply that any assumptions are made with respect to whether the language is an independent language or merely a variety of the lexifier language.

There has been much and heated debate of whether pidgins, creoles and mixed languages constitute a distinct type of language as opposed to languages that did not emerge due to intense contact. One of the starting points of this debate is the so-called Language Bioprogram Hypothesis formulated by Derek Bickerton (see, for example, Bickerton 1980 [1974]), which postulates that in the process of creolization the speakers will resort to the innate grammatical blueprint available to every human being, rather than carry over traits from the parent's language, which as a pidgin constituted an incomplete form of language input. One way of testing the hypothesis is to systematically compare contact languages with non-creoles[15] and see if they actually do pattern differently from non-creoles or if they are just treated as separate kinds of languages due to their socio-historical origin. This kind of testing is now made easier with large-scale databases. In an effort to test the validity of whether contact languages constitute their own type of language, I consistently compare the patterns found in Michaelis et al. (2013) with those found in Dryer & Haspelmath

15. Here and henceforth I use the expression 'non-creole(s)' as a short cut for meaning languages that are not classified as pidgins, creoles or mixed languages.

(2011) where possible in this book.[16] Needless to say, the patterns that emerge from the comparison between the two databases, whether similar or different, do not in any way imply any value judgement about the languages that make up the databases. The value of a linguistic system remains the same irrespective of whether it contains or lacks a given feature. For example, to state the fact that Mandarin has tone while English lacks that linguistic feature is merely an observation of the state of affairs and does not in any way imply that English is somehow a 'worse' or 'inferior' or in any way a lesser value language than Mandarin. This is equally true for comparisons between contact languages and non-creoles as well as for comparisons between spoken and signed languages (see below).

It should be noted that it is not necessarily the case that pidgins and creoles will display the same kinds of patterns, given their different domains of use. However, I have simplified matters considerably by treating the contact languages in Michaelis et al. (2013) as a single group of languages when comparing them with non-creoles. I stress that this simplification is a compromise made necessary due to space limitations. For more on why pidgins and creoles should be treated differently, see, for example, DeGraff (2009) with further references.

2.2.2 A brief note on sign languages

Sign languages are languages where the communication is achieved not by way of using the lungs, vocal cords and oral/nasal cavities to produce sounds, but by way of using the hands, upper body and face to produce signs. Sign languages are primarily, but not exclusively, used by deaf and hard of hearing communities across the world. They tend to be minority languages in the countries where they occur, and were, until very recently, to a large extent unrecognized or even repressed languages. For a long time sign languages were simply assumed to be sporadic or haphazard gestures. This is not the case at all. Sign languages represent as sophisticated linguistic systems as spoken languages do. The difference is simply that they are signed and not spoken. For an introductory overview of sign languages, see Sandler (2006) with further references. Johnston & Schembri (2009), Valli et al. (2005) and Sutton-Spence & Woll (1999) are very accessible introductions to specific sign languages (Australian Sign Language (Auslan) in Australia, American Sign Language (ASL) in the USA and British Sign Language (BSL) in the UK respectively), but also provide much general information on sign languages and sign language linguistics.

Systematic linguistic research on sign languages only started some 50 years ago, with the brilliant and highly influential *Sign Language Structure* by William Stokoe,

---

16. I am grateful to Magnus Huber for communicating to me the figures of the relevant features in the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013) according to their status on 27 March 2012.

published in 1960.[17] Since then a wealth of research has been produced, mainly on sign languages in North America and Western Europe. Sign languages occur all over the world, but the majority are still poorly described or not described at all. For this reason genealogical classifications of sign languages are not properly understood yet. In fact, it is not even clear if family trees in the traditional sense are relevant for sign languages. Most of the known links between sign languages today are based on the fact that they are young languages that often emerged through special education systems for the deaf.

> In a number of cases, special education for the deaf was first brought to a country from abroad, including the importation of some influence from the sign language of that other country. This potentially results in creolization of pre-existing indigenous forms of sign languages, sign pidgins, or so-called "home sign" systems (the latter two being less developed forms of gestural communication) with a foreign sign language. For example, American Sign Language is believed to have arisen from a creolization situation involving French Sign Language and pre-existing local sign varieties. While relationships between sign languages can thus be posited on the basis of historical knowledge, it is not clear whether these relationships can be considered "genealogical" in the same sense of the term as it is applied to spoken languages. (Zeshan 2011c)

The following groupings of sign languages have been proposed (from Zeshan 2011c with minor modifications):

BSL – Auslan – New Zealand Sign Language (NZSL: New Zealand)

Japanese Sign Language (NS: Japan) – Taiwanese Sign Language (TZS: Taiwan) – South Korean Sign Language (SKSL: South Korea)

French Sign Language (LSF: France) – ASL – Russian Sign Language (RZY: Russia) – Dutch Sign Language (NGT: the Netherlands) – Flemish Sign Language (VGT: Belgium) – Quebec Sign Language (LSQ: Canada) – Irish Sign Language (TCE: Ireland) – Brazilian Sign Language (LSB: Brazil)

DGS – (perhaps also other sign languages in Europe and the Middle East) – Israeli Sign Language (ISL: Israel)

Swedish Sign Language (ST: Sweden) – Finnish Sign Language (SV: Finland)

ASL – Ugandan Sign Language (USL: Uganda) – Thai Sign Language (TSL: Thailand) – Kenyan Sign Language (KSL: Kenya)

Italian Sign Language (LIS: Italy) – Argentine Sign Language (LSA: Argentina)

Chinese Sign Language (CSL: China) – Hong Kong Sign Language (HKSL: China)

---

17. In fact Barnard T. M. Tervoort published an even earlier systematic study of sign language as a linguistic system with his *Structurele analyse van visueel taalgebruik binnen een groep dove kinderen* (*Structural Analysis of Visual Language Use in a Group of Deaf Children*) (1953). Presumably the impact was slightly less than Stokoe's (1960) publication because fewer people can read Dutch.

It should be noted that sign languages may also arise on their own, without being spurred by educational programs. Adamorobe Sign Language (AdaSL) in Ghana is an example of such a language (cf. Nyst 2007), as is Urubú Sign Language (UKSL) in Brazil (Kakumasu 1968) and Kata Kolok (KK) in Bali (Branson et al. 1996).

It is important to keep in mind that sign language names mostly simply indicate where the language is spoken, as in *British Sign Language*, which is used in Great Britain, or *Deutsche Gebärdensprache* (German Sign Language), which is used in Germany. This should not be taken to indicate that they are signed versions of the spoken language of the area. They are not. They are languages in their own right and can be quite different from the spoken languages of the society they are located in. DGS, for example, has subject-object-verb word order, while spoken German has subject-verb-object word order. Sign languages are, however, in intense contact situations with the spoken languages of the area. The absolute majority of signers must also be able to function in the spoken language of the society they live in, if nothing else in written form. For more on language contact between spoken and signed languages, see 13.3.

Sign languages constitute their own type by virtue of the difference in modality (using visual-gestural communication instead of audio-oral communication). Including sign languages systematically in typological surveys and research will allow us to investigate whether or not this difference in modality actually entails differences in language structure. For this reason, and as a general appeal to include sign languages systematically in cross-linguistic studies, a section on sign language has been included in all the chapters dealing with typological data in this book. Since vastly more data is available for spoken languages, the sections on sign languages are by necessity shorter than the sections on spoken languages. Future documentation of and research on sign languages is eagerly awaited and will provide invaluable information on the nature of language as a human phenomenon.

2.3 What are universals?

Cross-linguistic surveys allow us to study patterns that systematically occur across languages. Recurring patterns allow us to make typological generalizations and formulate **language universals**. Language universals refer to properties that hold for all or most known human languages. It is important here to keep in mind that the term language universals, as used in typology, refers to quantitative statements that are based on cross-linguistic studies. Or to be more exact:

> Typological universals are empirically established generalizations that describe distributional patterns for particular grammatical phenomena across languages. These distributional patterns are regarded as universals to the extent that they are found in all languages or in a statistically significant number of languages. (Cristofaro 2010: 227)

The term 'universal' is also used by Generative Grammar to denote a feature common to all human languages, but there the claim is not based on quantitative studies.

Typology differentiates between different kinds of universals, which will be briefly defined and summarized below. For more on language universals, see, for example, Moravcsik (2010) with further references, which this section relies on a great deal.

2.3.1 Unrestricted universals: absolute and statistical

With **unrestricted universals** we make statements about independent phenomena in languages without any further conditions to those phenomena. That is, we simply state that X is present in all or most of the known languages. Universals can be of two types, absolute or statistical. **Absolute universals** are universals that hold for every single human language, without exceptions; the assumption is that the feature must be present in any and all languages. **Statistical** (or **probabilistic**) universals hold for most, but not all, languages; the assumption is that the feature is likely to be present in a language.[18] Examples of absolute universals are:

(9) a. All spoken languages have vowels
b. All languages can refer to entities
c. All languages have ways of forming questions

The statements in (9) can safely be assumed to hold without exceptions. While (9a) is not applicable to sign languages (but see 4.3 for an overview of sign language phonology), (9b and c) hold for both spoken and signed languages. Examples of statistical universals are:

(10) a. Most spoken languages have the nasal /n/ (but not Central Rotokas (West Bougainville (West Bougainville): Papua New Guinea), which lacks nasals altogether; (Robinson 2006: 207))
b. In most languages the singular is the base form and the plural is the overtly marked form (but not Aari (Afro-Asiatic (South Omotic): Ethiopia), which has no plural but an overtly marked singulative; (Hayward 1990: 444))[19]

---

18. Strictly speaking a statistical (or nonabsolute) universal is a contradiction in terms and should, more accurately, be called 'statistical tendencies'. It has, however, become a convention to distinguish between absolute and statistical universals in linguistic typology. I will follow this tradition throughout this book.

19. The terms 'marked' and 'markedness' tend to be used with a variety of meanings. In this book I use it to mean 'indicated'. An overt marker is thus an overt indication of something, i.e. a form indicating some kind of meaning. For a discussion on the various usages of the term 'marked' (and suggestions for why the term should be abandoned), see Haspelmath (2006).

(10) c. Most spoken languages employ a rising intonation for yes-no questions (but not Hawai'i Creole English, where yes-no questions have falling intonation; based on own fieldwork)

The statements in (10) hold true for an overwhelming majority of languages in the world. (10a and c) only hold for spoken languages, while (10b) holds for both spoken and signed languages.

As mentioned before, it is simply not possible to include all human languages in a survey. Therefore it is important to keep in mind that all universals are hypotheses. Even if a feature is present in all languages investigated, and thus counts as absolute, there is always the possibility that new data will reveal new systems and provide exceptions to the universals formulated. It is then also vital that this data be made widely available. For example, until 1977 it was commonly believed that there was no evidence for a language with an object initial word order (object-verb-subject or object-subject-verb), so it was argued that object initial word order was impossible for human language. These assumptions were made despite the fact that there actually had been reports of languages with object initial word order, such as Beauvoir's (1915) and Tonelli's (1926) descriptions of Selknam (Chon (Chon Proper): Argentina).[20] But only with Derbyshire's publication in 1977 on the word order of Hixkaryana was this absolute universal widely accepted to have been proven wrong.

Hixkaryana (Cariban (Cariban): Brazil)

(11)

yahutxho	matkahekonà	wosà
manioc.peel	she.was.pounding.it	woman
Object	Verb	Subject

'A woman was pounding manioc peel.' (Derbyshire 1977: 597)

In (11) the object (*yahutxho*) precedes the verb (*matkahekonà*), which in turn precedes the subject (*wosà*). Since then object initial word order has been found in a number of other languages (for more details, Chapter 10). This, in a sense, shifted the universal from an absolute to a statistical one. Despite the exceptions to the universal, it is, to our knowledge, still rare to have the object clause initially. Thus a statistical universal that hypothesizes that languages avoid object initial word order has not been rendered invalid. It is simply not an absolute anymore.

While absolute universals need only one exception to be falsified, statistical universals are also possible to falsify, although that demands more data. For example, a common claim in typology has been that languages that have object-verb word order also tend to have adjective-noun word order, as in Ainu:

---

20. I am grateful to Harald Hammarström for bringing these sources to my attention.

Ainu (Isolate: Japan)

(12) a. aynu kamuy rayke
person bear kill
Subject Object Verb
'The man killed the bear.'

b. pirka kewtum
good heart
Adjective Noun
'good heart' (Shibatani 1990: 23)

In (12a) the object (*kamuy*) precedes the verb (*rayke*) and in (12b) the adjective (*pirka*) precedes the noun (*kewtum*). However, when investigating a very large sample of 1316 languages, Dryer (2011x) found no significant correlation between the order of the object and the verb and the order of the adjective and the noun. Statistical universals are thus falsifiable, but it demands access to a lot of data. Hence both kinds of predictions are testable and falsifiable, one of the main criteria for a scientifically viable hypothesis.

2.3.2 Implicational universals

With **implicational** (or **restricted**, also called **typological**) **universals** we have preconditions to the universal and make statements of the "if X, then Y" kind. That is, we hypothesize about correlations between features. Implicational universals may also be absolute or statistical. Examples of absolute implicational universals are:

(13) a. If a language has the phoneme /t/ then it also has the phoneme /k/ (Pericliev 2008: 206)

b. If a language has reflexives for the first and second person, it will also have reflexives for the third person (Comrie 1989: 19)

In (13) the hypothesis is that the implications hold for all languages. Examples of statistical implicational universals are:

(14) a. If a language has the phoneme /n/ it is also likely to have the phoneme /m/ (but not Konkani (Indo-European (Indic): India), which has /n/ but no /m/; UPSID:[21] *sv* Konkani)

b. If a language has object-verb word order, it is also likely to have postpositions (but not Persian (Indo-European (Iranian): Iran), which has object-verb word order but prepositions; (Dryer 2011y))

---

21. All data from UPSID (UCLA Phonological Segment Inventory Database) is based on the interface set up by Henning Reetz (http://web.phonetik.uni-frankfurt.de/upsid_info.html [accessed in March 2010]).

In (14) the hypothesis is that the predicted implications hold for most languages.

Implicational universals can be either one-way or two-way predictions. A **bidirectional implicational universal** is a prediction that works two ways. What this means is that we can hypothesize that if a language has X, then it also has Y, and conversely, if it has Y, then it also has X. An example of a bidirectional implicational universal is (14b); a language that has object-verb word order also tends to have postpositions. That means that if we see a language with object-verb clausal word order, we can expect it to have postposition. But it also means that if we see a language with postpositions, we can expect it to have object-verb word order. Furthermore, we can reverse the prediction, and say that if a language has the opposite kind of word order, namely verb-object, then it is also likely to have prepositions. And conversely, if a language has prepositions, then it is also likely to have verb-object word order. The correlation works both ways.

Not all implicational universals are bidirectional. A **unidirectional implicational universal** is a hypothesis that only holds one way. An example of a unidirectional universal is

(15) If in a language the relative clause precedes the noun, then it is usually has an object-verb word order while if a language has verb-object word order, then the relative clause usually follows the noun (Dryer 2011z)

The prediction in (15) only holds as a one way prediction. While it is possible to say that most of those languages where the relative clause precedes the noun also have an object-verb word order, it is not possible to say that if a language has object-verb word order, then the relative clause will precede the noun, because there are many languages with object-verb word order where the relative clause follows the noun. Nor is it possible to reverse the prediction and say that if in a language the relative clause follows the noun, then it will also have verb-object word order. The implication only holds one way.

2.4 A very short note on motivations for language universals

It is beyond the scope of this book to provide a discussion on the motivations for the various cross-linguistic patterns we see, something that could in itself fill several volumes. This section makes no claim to do anything more than presenting a few key concepts central to the debate on explanations for linguistic universals. For an exceptionally accessible discussion on the motivations for language universals, see Whaley (1997: 43ff). See also Croft (2003: 49ff), Haiman (2010) and Moravcsik (2010) with further references, as well as Croft (2002) for a syntactico-semantic model for explaining language universals.

It is a matter of debate whether explanations for language universals should be sought language internally or externally. Language **internal** explanations are based on the structural system of the language in question, while language **external** explanations are based on factors outside the structural system of the language.

An example of a language internal explanation is the notion of **iconicity**, with the principle that the formal expressions in a language express semantic notions. By this notion the more complex the form is, the more complex is the notion that it represents. For instance, *book* (one item) is a simpler notion than *books* (several of the item in question), which is also reflected in the difference in morphology, where the simpler notion has the simpler form and the more complex notion has an extended form (plural *-s*). For more on iconicity and markedness, see, for example Hawkins (2001) and Bybee (2010).

Examples of external explanations for language universals are the roles of discourse, processing and economy. **Discourse** refers to a connected series of utterances by speakers. Humans structure their speech in order to convey a coherent message. The motivation to form a cognitively efficient discourse will shape the structure of languages. In a passage like *Jenny saw the man. She thought she had seen him before…* the pronouns (the referents) come after the nouns (the antecedents), which makes for a more coherent message than something like *She saw him. Jenny thought she had seen the man before…*, where the referents precede the antecedents. For more on discourse and language universals, see Kärkkäinen et al. (2007) and Hopper & Thompson (1993) with further references.

Economy refers to two processes: (*i*) frequently used elements tend to get reduced, or, put differently, the length of a word correlates with how often it is used; and (*ii*) elements that are highly predictable tend to get eliminated. An example of (*i*) is that the shorter *and* is more a common word than *before*, which in turn is more common than *cardboard*. An example of (*ii*) is a sentence like *When John entered the room Peter left*, where *the room* does not need to be repeated because the context implies what it was Peter left (so-called ellipsis). A classic on economy in relation to iconicity is Haiman (1983).

Processing refers to the cognitive effort it takes to comprehend linguistic structures. Here the motivation for language universals is argued to be that those linguistic structures that are easy to process are preferred. A sentence like *Mary looked the reference John asked for up* takes more effort to process than *Mary looked up the reference John asked for*. The latter structure is therefore expected to be preferred. For more on processing, see, for example, Hawkins (2001 and 2010) with further references.

2.5 Summary

Typology concerns itself with the differences and similarities between linguistic systems, both within and across languages, and investigates patterns of distributions of linguistic structures, something which demands cross-linguistic surveys. Comparison of linguistic types may be either synchronic or diachronic. Any part of the linguistic system may serve as a starting point for typological comparisons. With its cross-linguistic surveys, typology may help show what kinds of recurring patterns can be found, and as such may serve as a tool for testing hypotheses of language and linguistic systems.

Classification is a central concept in linguistic typology, especially the genealogical affiliation, primarily according to the family and genus. Languages are assumed to descend from one ancestor and to have branched out from that ancestral root. The family is the highest level of classification. The genus is a genealogical sublevel of classification that is intended to be comparable across the world with regard to time depth. Since languages that are in contact with each other are likely to influence each other in some way, the area in which languages are spoken is an equally important classificatory parameter for typological surveys.

Due to their socio-historical background, pidgins, creoles and mixed languages are not easily classified genealogically. These contact languages have in common that they emerged in situations of intense contact. Pidgins are reduced languages used in situations where there is no common language between the parties and are typically not native languages to anyone. Creoles typically emerge from pidgins and are native languages of entire communities. Mixed languages typically emerge from stable bilingual situations.

Sign languages were not recognized as full linguistic systems on par with spoken languages until the 1950s and systematic linguistic research on sign languages has only been conducted since the 1960s. As yet there is not enough data for genealogical classifications of sign languages. Sign languages may either emerge spontaneously or in controlled environments due to education systems.

Universals are typological generalizations based on cross-linguistic surveys. Absolute universals hypothesize that a linguistic phenomenon is found (or is lacking) in all human languages. Statistical universals hypothesize that a linguistic phenomenon is found (or lacking) in most human languages. Implicational universals are hypotheses of the "if X, then Y" kind and can be absolute or statistical. Implicational universals can be bidirectional, where the implication can be reversed, as in "if X, then Y" and "if Y, then X", or they can be unidirectional, where only "if X, then Y" holds but where the reversed "if Y, then X" does not hold.

There may be various motivations for language universals, both internal, such as iconicity, and external, such as cognitive processing, economy or coherent discourse.

2.6 Keywords

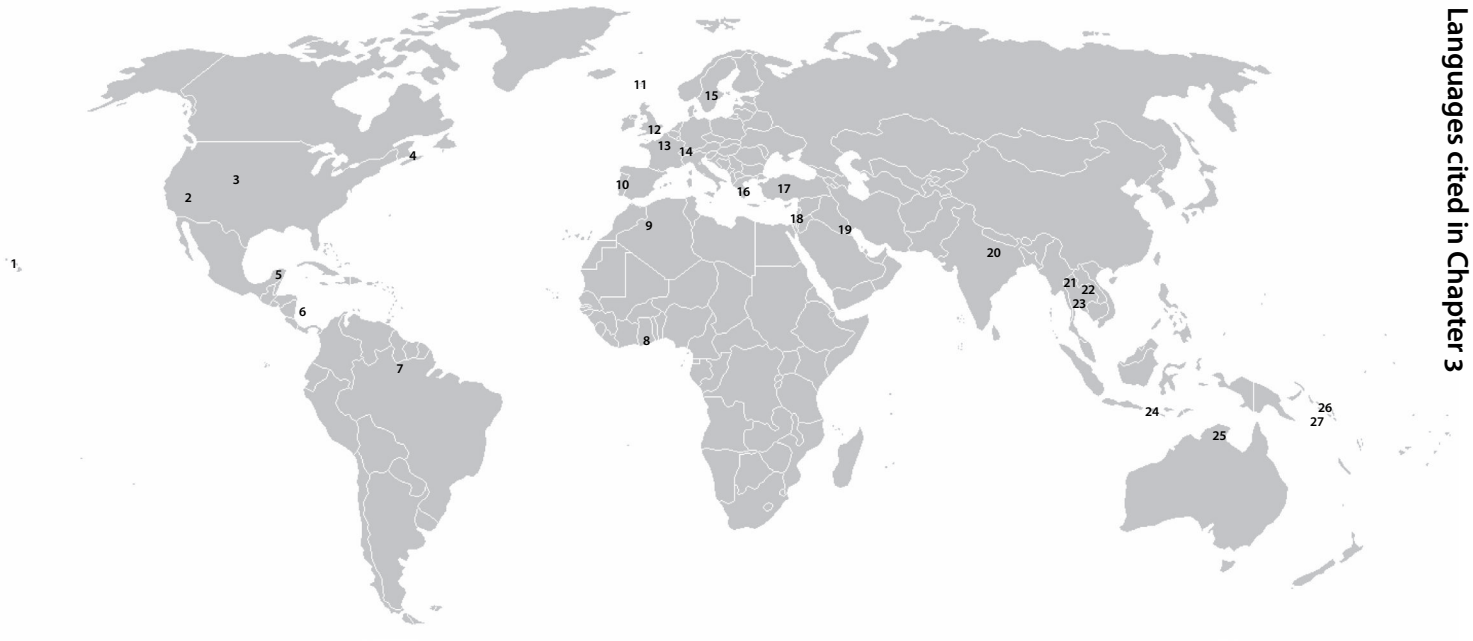
absolute/statistical universals
classification
creole
family
genus
language isolate
mixed language
motivations for universals
pidgin
sign language
unidirectional/bidirectional universals
unrestricted/implicational universals

2.7 Exercises

1. Define **family**, **genus** and **language isolate**.
2. Why do pidgin and creole languages get treated as a distinct type of language?
3. Why does typology demand cross-linguistic mapping?
4. Is it justified to treat sign languages as a distinct type of language? Motivate your answer.
5. Is the following statement true or false? Motivate your answer.

   Statistical universals are not falsifiable and are therefore less valuable than absolute universals for understanding human language.

Languages cited in Chapter 3



1 Hawai'i Pidgin Sign Language
2 Navajo
3 Plains-Indian Sign Language
4 Maritime Sign Language
5 Yucatec
6 Providencia Sign Language
7 Hixkaryana
8 Adamorobe Sign Language
Ghanaian Sign Language
9 Berber, Figuig
10 Portuguese
11 Faroese
12 English
13 French
14 Alemannic, Swiss Romansch
15 Swedish
16 Greek, Ancient
17 Turkish
18 Hebrew, Biblical
19 Sumerian
20 Sanskrit
21 Old Chiangmai Sign Language
22 Ban Khor Sign Language
23 Old Bangkok Sign Language
Thai Sign Language
24 Kata Kolok
25 Nunggubuyu
26 Lavukaleve
27 Rennellese Sign Language

Chapter 3

Methodology

Sampling, databases, and how many languages does a typologist speak?

All linguistic research is dependent on data in some form. For large-scale typological surveys access to data is of paramount importance. This chapter discusses the various kinds of sources for language data (3.1), with a note on the situation of the languages of the world (3.1.1) as well as a section on fieldwork, documentation and description of languages (3.1.2). Section 3.2 brings up the issue of sampling, including a brief discussion on types of language samples (3.2.1) and types of bias (3.2.2). In Section 3.3 I give examples of three rather different kinds of databases. Section 3.4 mentions some methodological issues related to sign language typology.

3.1 Data

Figure 3.1 Idea: Magnus Huber. Illustration: Patrick Thornhill. Used with permission.

Perhaps the most off-turning question that a linguist has to keep swatting away like an insistent mosquito is "how many languages do you speak?". It is a question that I, personally, absolutely refuse to answer, partly due to the fact that there are certain times of day when I feel I don't speak any languages at all, but mostly due to the implicit but unwarranted assumption that a person can only do research in languages that he or she personally speaks. The implication, then, would be that all languages for which we have data but that have gone extinct, such as Ancient Greek, Sumerian (Isolate: present-day Iraq), Sanskrit (Indo-European (Indic): present-day India), Biblical Hebrew (Afro-Asiatic (Semitic): present-day Israel), and so on, would be unavailable for linguistic research because there are no speakers left.

All linguistic research is dependent on language data, and the more data is available, the more nuanced the investigation of it can be. However, data does not only come in the form of a native or native-like knowledge by someone who also happens to be a trained linguist. Furthermore, as was shown in the previous chapter, much typological research is dependent on data from a high number of languages. Needless to say, a typologist does not personally know all the languages that are included in a database. (Oddly enough, a recurring question I have had even from other linguists is precisely along the lines of "how well can you know hundreds of languages?".) For example, I made four separate surveys of various linguistic features while writing this book (see 6.2.1.3, 8.2, 9.2 and 11.2.3.1), which, in total, involved 526 languages. I did not learn 526 languages during the course of writing this book. What all linguists have to do is to combine different sources for data. Apart from native speaker intuition and expert knowledge, linguistic data derives from descriptions (primarily grammars), elicitation and texts. Each of these sources has its merits and shortcomings, but used together they can provide a rather detailed picture of a language. All of these sources ultimately derive from various kinds of field- or archival work.

Language descriptions, especially **descriptive grammars**, are one of the main data sources for typological research. The major advantage of descriptive grammars is that they typically are comprehensive analyses of a language, done by either native speakers or fieldworkers (who tend to become experts on the languages they do fieldwork in) that are also trained linguists. Most descriptive grammars are organized according to rather similar principles, which means that they are to a large extent cross-compatible. However, no language description is without its theoretical bias; no matter how hard the linguist tries to be neutral in his or her description, theoretical premises will sneak in. Such things as the way to go about analysing a language, how to get started, how to structure the analysis, etc., all ultimately rely on one's theoretical background. Language attitudes and the socio-political context of the language may also affect the orientation of the grammar. For example, a native speaker of a stigmatized language may unconsciously be biased towards proving the worth of the language

by establishing that it has certain kinds of features. Whether it is due to language attitudes or theoretical orientation, there is always a risk that the description seeks to answer such questions as "what does feature X or Y look like in this language?" or "is this an instance of A or B" (cf. Gil 2001). The problem with such questions is that they presuppose the existence of feature X or Y or that A or B are relevant categories in the language in question. But languages differ and, as we saw in the previous chapter, what is a relevant category for one language may not be at all relevant for another. Try to imagine what the result would be if someone set out to describe the tone system of English, approaching it with the question "which tones does this language have?" and "do the tones of English have any semantic patterning?". Since English does not have tone, that would be a rather nonsensical endeavour.

Descriptive grammars may of course also be of varying quality and provide varying levels of details. Furthermore, a grammar is by necessity an approximation: first of all, it is impossible to capture all facets of a language; secondly, for a grammar to be publishable it cannot be of infinite length. Compromises have to be made. Therefore very specific questions may not be answered in a descriptive grammar. Information for such specialized investigations as my survey on expressions for 'It is raining' (9.2), for example, may or may not be available in a description (for example, of the ca 800 sources I literally looked through every page of, I found the necessary information for only about a quarter of the languages).

The way to go about finding data in a grammar of a language under investigation is not only to read what the expert states about the language, but also to look at the examples provided. Ideally the examples are given with morpheme-by-morpheme glossing (see 1.3.2) and translation. When doing the survey of expressions for 'It is raining', for example, I was mostly dependent on finding an example involving the expression, since few grammars actually had a section discussing weather expressions. Thus, when coming across an example like (16) below I was able to establish which kind of strategy the language had:

Figuig Berber (Afro-Asiatic (Berber): Morocco)

(16) t-ttaɣ t-bica
3SG.F-make.INT F-rain
V S
'It is raining.' (Kossmann 1997: 249)

In (16) I used the translation to identify the expression I was looking for. I then used the interlinear glossing to analyse how the expression was constructed. I copied the example, including the glossing, translation and what the abbreviations meant, and noted the bibliographical details in a file where I collected all the data. I also assigned a value in a spreadsheet for the language based on the information I had found. This

procedure was then repeated for each language that I investigated (as mentioned above, however, only about a quarter of the sources I looked through actually had an example that I was able to use).

An important consideration when doing typological surveys based on descriptive grammars and similar kinds of sources is that the author of the description may use terminology or define categories differently from the linguist doing the survey. It is also important to keep in mind that the reader of the grammar is limited by the fact that s/he is not likely to have firsthand knowledge of the language described. Each of us doing a survey must therefore make every effort to understand exactly what the author of a description means with his or her terms and categories. The combined information of what the author of the description means and the examples given in the description will allow the typologist to decide which feature value seems most appropriate for the language. For example, the typologist may have a certain definition for a linguistic feature, such as that 'perfect' should be used both resultatively and experientially (as in Dahl & Velupillai 2011d). If a language employs what the description calls 'perfect' for only one of these uses, then the language will be coded as not having 'perfect' for the survey in question, since it did not conform to the set criteria of the investigation. The survey and the description may thus differ in their respective analyses; what is crucial for a survey, however, is that the definitions of the feature values that are being investigated are applied consistently for all languages.

Another source of data is **elicitation**. With elicitation the linguist asks native speakers or experts with native-like competence pointed questions to glean information about a language. Elicitation thus allows the linguist to ask about the particular features s/he is investigating. However, apart from the fact that elicitation is extremely time-consuming for both the linguist and the **language consultant** (also called **informant**), which makes it an impractical source of data for hundreds of languages, it is also problematic due to the fact that it is not a natural language situation. The native speaker gets questions that are isolated and pretty much out of context, and the answers given, i.e. the data received, are likely to reflect this rather stilted situation. We have all experienced that repeating an ordinary word or a sentence many times can make it sound very odd (or, conversely, that repeating a nonsense word or sentence many times can make it sound perfectly normal). Added to this is the so-called **Observer's Paradox**, a term coined by William Labov (1972) to capture the paradoxical problem that the mere presence of a linguist will affect the language s/he is investigating, since the mere presence of the linguist will make the native speaker more self-conscious of his or her language, which will affect the naturalness of the language. In elicitation situations the native speaker is likely to be highly conscious of his or her language. Furthermore, elicitation depends on the questions the linguist can think to ask, which means that patterns and features that the linguist is not aware of are not likely to be

captured, since the linguist didn't know to ask for that particular information. The linguist's prior expectation thus shapes the information gleaned which consequently makes the data selective. Another risk with elicitation is that the questions asked are not typically formed in the consultant's language, but have to be translated either by the language consultant or by an assistant knowledgeable in both languages. This may affect the data in various ways.

A different form of elicitation is to construct **questionnaires**, as, for example, Östen Dahl did (1985). This allows the linguist to conduct multiple parallel investigations. Questionnaires are essentially elicitation in written form that is further removed from the interaction between the linguist and native speaker. The various problems that elicitation situations bring with them thus also apply to questionnaires. Furthermore, designing questionnaires in such a way that misunderstandings are minimized is an almost impossible task. The risk is not only that the native speakers might misunderstand questions, but also that the linguist might misunderstand the answers. Another point is that questionnaires will typically be designed in a language other than that which the data is sought from. Last, but not least, it depends on getting hold of enough language consultants and experts that are both willing and capable of devoting the time and energy that filling out detailed questionnaires requires.

A third source of data is **texts**. With texts not only written records are meant, but also longer passages of spoken language that have been recorded and transcribed. The latter includes conversations, narratives, ritualistic monologues, and so on. This data source is perhaps the most time-consuming of all, so much so that the sources discussed above may seem as shortcuts in comparison (to quote Tim Thornes, p.c.). Collecting textual data demands dedicated field- or archival work; many linguists spend years and decades on the languages they investigate (for more on fieldwork, see below). These kinds of texts are usually come by through recordings in different settings, for instance conversations in familiar environments such as the workplace or the home of the native speaker. This means that the language is likely to be more naturalistic than in the artificial situation of elicitation. However, the Observer's Paradox is at play here too, even if at a lesser degree than in elicitation sessions, as the environment is much less contrived. Still, when recordings of naturalistic spoken language data are being made, the language consultants are likely to be aware of the presence of the linguist (unless unethical methods are employed, such as recording people without their permission. Data of this kind should be avoided). One way around this problem is to include native listeners in the recording session, or training community members in documentation practices.

The Observer's paradox would of course not play a role for written texts, but written and spoken language may differ: with written language you do not have the benefit of facing the addressee, which means that you have to phrase yourself in a manner that ensures

to the largest possible extent that your message gets across. This is likely to affect how you use your language. Furthermore, many languages lack a written tradition altogether.

With texts we usually have a larger amount of data, which allows for quantitative analyses. However, not all linguistic phenomena will necessarily emerge during a conversation or in a narrative, especially rare or specialized ones. Moreover, the linguist may or may not be able to properly interpret a text or to capture the various features and nuances through interlinear glossing. And again, the linguist is likely to be influenced by his or her theoretical background (cf. above). Analysing the text outside its original context may also hamper analyses and interpretations. Ideally, the linguist would be able to consult native speakers to cross-check that the interpretation is correct, but this is rarely possible, and is pretty much impossible when dealing with already extinct languages.

In sum, ideally one would wish to have access to all three data sources, grammars, texts and possibilities for elicitation. Unfortunately this is rarely possible, but "any and all sources can provide relevant data when used judiciously" (Croft 2003: 30).

3.1.1 A note on the situation of the languages of the world

There are roughly 7000 known languages in the world today (Lewis 2009). Of these, fewer than a third have been adequately described (cf. Payne 1997, Romaine 2007, with further references). By adequately described I mean that a full descriptive grammar and preferably also a dictionary has been produced on the language. The remaining (more than 70%) languages of the world have either only received sporadic attention from linguists, or none at all, although intense efforts are being made by linguists all over the world to try and change this. We thus lack written and recorded data for the majority of the existing languages in the world. At the same time, languages are disappearing at an alarming rate. While absolute figures are not possible to come by, due to, for example, uneven distributions of census material, it is reasonable to assume that at the rate things are going now, roughly half of the languages of the world will be extinct by the end of this century (cf. Krauss 1992 and 2007).[22] This is a state of much higher crisis than the already alarming state of endangerment of biodiversity. And just as extinction of plant and animal species is a threat to the well-being of our world, so is the extinction of languages (cf., for example, Romaine 2007 and Hale 1992 for discussions). With the death of a language, a host of human knowledge is lost. A language is more than a tool for communication. It is not only the breath of a culture, but also embodies the knowledge accumulated throughout the generations of that culture. Specific ways of solving problems, of maintaining an ecobalance in the traditional

22. Some predict a loss of 60–90% of the currently known languages within the next 100 years (Romaine 2007: 115 with further references).

territory, and of interacting with the world, from mathematical skills and perspectives to agricultural tips and tricks and anything in between, will become embodied in a language as it grows and develops with the community that speaks it. Language extinction is thus not merely a loss of opportunity for linguists to study diverse linguistic systems, but also a loss of valuable human knowledge in general. For a very accessible discussion of what the world loses when languages die, see Harrison (2007).

How to determine whether a language is endangered or not is not necessarily straightforward. A "safe" or **stable** language typically has official state support, is used in schools for teaching, is used in the media, and is probably also used in written communication. It probably has a rather sizeable population of speakers. An **endangered** language has a dwindling number of speakers, with fewer and fewer speakers transmitting the language to the next generation (that is, fewer and fewer speakers learn the language natively). Note, however, that size alone is not a guarantee for the stability of a language. Navajo (Na-Dene (Athapaskan): USA), for example, has almost 150,000 speakers, while Faroese, for example, has fewer than 50,000 speakers (Lewis 2009). Faroese is at this stage not endangered: most Faroese children learn Faroese as their mother tongue, use Faroese in school, write texts in Faroese, encounter Faroese in the media, and grow up to speak Faroese to their own children. This is not the case with Navajo. In 1968 about 90% of the first grade children had Navajo as their mother tongue. In 1998, only one generation later, this number had dwindled to 30% of the first grade children. What this means is that fewer and fewer children learn Navajo as their mother tongue. Eventually we might have a situation where nobody learns Navajo natively. That would make it a **moribund**[23] language, i.e. a language destined to go extinct when the last generation of speakers dies. For overviews of endangered languages, see, among others, the chapters in Brenzinger (2008) and Miyaoka et al. (2007). For a discussion on the causes of language endangerment and the effects of language policies on endangered languages, see the chapters in Derhemi (2002).

There are many and varied reasons for language extinction, and in many cases they are situation specific. The most dramatic reason for the world having lost languages is genocide, the mass extinction of an entire culture. Other reasons include displacements of a community, such as more or less forced urbanization due to socio-economical factors. For example, with the destruction of the environment due to pollution or deforestation, people might find themselves unable to lead a sustainable life in their traditional territories, and decide to migrate to other places, usually urban centres.[24] With that they might have to adopt new languages in order to get by in the new environment.

---

23. These kinds of languages are often also called endangered, e.g. in *Ethnologue* (Lewis 2009).

24. In fact, this is another reason why we should sit up and take notice of the rate of language death in the world: the causes of language death and language endangerment can actually be seen as an

They might further feel that the new language would give their children better opportunities for the future, opting therefore to not pass on their own native language to their children. If this is done on a large scale, the community might end up shifting to another language altogether and losing the original language. Another reason for language endangerment is discrimination, such as forced monolingualism, which usually derives from the power holder's wish for uniformity (which is usually easier to control). This can take the shape of forced assimilation, where a language is actively suppressed, for example by meting out punishment to those who use it, or by assimilatory education, where only the preferred languages are taught and promoted.

It should be pointed out here that **language shift**, which in turn carries with it language extinction (because the speakers shift from one language to another, usually from a smaller language to a dominating language), may come about as a voluntary act by speakers. Very often this is because shifting to another language is expected to yield better opportunities for the next generation. As Ladefoged (1992) points out, as linguists we must respect speakers' choices in what language they wish to speak and not presume to know what is best for a community. My personal take on the situation is that any kind of repression, among which I count forced monolingualism and forced language shift (not to mention destruction of territory and forced displacement), should be abandoned. While it is understandable that political administrations need a certain degree of unity, there is no reason to suppose that multilingualism threatens unity and stability. People are perfectly capable of handling more than one language and can easily function in both an official administrative language and in other languages at the same time. It seems to me that people should be free to use whichever languages they like (and choose where and how they want to live); forced streamlining is not helpful, while voluntary diversity is enriching both culturally and scientifically.

3.1.2 A note on language documentation and description

Language documentation and description serves a number of purposes. By documenting languages we engage in amassing data for preservation. This will allow future generations to access data for languages even after they are gone. Describing languages provides information about them, and has the extended effect that the mere existence of the descriptions may be empowering for endangered languages. Both of these endeavours are dependent on fieldwork of various kinds.

Fieldwork is essentially the procedure of acquiring linguistic data from language consultants, preferably in environments familiar to them, such as their homes or workplaces. This can mean anything from interviewing a colleague sitting in the office next

---

indicator of ecological destruction, and thus a sign of a collapse of our global ecosystem, as pointed out in Nettle & Romaine (2002).

to you, to going to some small community in a remote place to interview speakers from a small, possibly endangered, language, or anything in between. The latter type of fieldwork is what Hyman (2001) calls prototypical[25] fieldwork, and is probably what most readily springs to mind when hearing the term: a linguist spending an extended amount of time with a community in an exotic place, documenting and recording a little known language of a community with the help of local informants. But fieldwork can also be carried out in one's hometown, for example investigating the language of a specific community in the city or investigating different varieties of the same language in one place along various sociolinguistic parameters. Going to investigate the language of the village your grandfather was born in is also fieldwork. A good general definition of fieldwork is that it "describes the activity of a researcher systematically analysing parts of a language, usually other than one's native language and usually within a community of speakers of that language" (Sakel & Everett 2012: 5).

Irrespective of where fieldwork is conducted, it involves a multitude of considerations. First of all, people have to be willing to participate, to function as consultants. Their consent to participate must be informed, that is, they must understand what the whole endeavour is about. Many of the issues have to be considered individually for different cultures. The issue of compensation, for example, can be sensitive in various cultures. When planning for fieldwork, whether the prototypical kind or not, the linguist will have to consider not only what s/he wishes to investigate, but how s/he will be accepted into the community, how s/he will work with the consultants, and how s/he will go about gathering data, how to make that data as reliable and comprehensive as possible, and so on. A host of practical and logistical issues need to be solved, such as getting to the place, staying there, getting permission to work with human subjects, what kinds of equipment to use, and so on. For exceptionally accessible introductions to fieldwork for beginners, see Crowley (2007), which mainly focuses on the prototypical kind of fieldwork, and Sakel & Everett (2012), which gives guidelines also for other kinds of fieldwork.

Fieldwork allows us to carry out two very important separate yet interrelated tasks: to document a language and to describe it. **Language documentation** essentially means the collection of raw data that may then be used for further analysis. This means that the investigator collects field notes, makes recordings of naturalistic speech, conducts interviews and elicitations, records and observes participants, and so on. To make the data available it will have to be transcribed and translated, which

---

25. The term **prototypical** essentially means 'most typical example'. For instance, a prototypical bird is an animal with wings, two feet, a beak and feathers; it lays eggs and can fly. Some birds conform to all these values, such as robins and seagulls. Other birds do not completely conform to the typical example; penguins, for example, do not fly. A robin is thus a prototypical example of a bird, while a penguin is not.

by necessity involves some analysis of it. But by and large the purpose of language documentation is simply to preserve representative samples of the language, preferably a sizeable amount, so that specialists (both linguists, anthropologists, sociologists, historians, etc.) can work on the language, but also so that the language community itself has a record of its language for posteriority. The merits of language documentation is thus not only to give linguists raw data to work with, but also to preserve a cultural heritage of the community. Language documentation may also serve as the basis for awareness and revitalization programs, by, for example, serving as a base for teaching materials (see below).

While language documentation may not slow the rate of language death, it will, at the very least, preserve records of the languages for future generations. An example of a large scale language documentation program is *The Rosetta Project* (http://rosettaproject.org/), where information for over 1500 languages is currently stored, much of it publicly available. In principle this can be thought of as a kind of museum for languages. A special feature of the Rosetta Project is that the information for each language follows a set format, so that the information for every language parallels that of other languages. By having the same kinds of texts and set of vocabulary, the information is highly compatible for cross-linguistic comparison, which makes it an invaluable tool for typological research. For a wealth of information about language documentation and how to go about it, see the chapters in Gippert et al. (2006) and the further links at the official webpage of the book (http://titus.uni-frankfurt.de/ld/). The open access journal *Language Documentation and Conservation* (http://nflrc.hawaii.edu/ldc/) publishes articles on a wide variety of topics relating to language documentation.

Language description is a separate, yet interrelated, task from language documentation. A description seeks to illustrate the essentials of a language, based on available material. It provides analyses of a variety of areas of the language, such as its phonological, morphological, grammatical and syntactic systems, as well as, ideally, presenting a lexicon of the language. Ideally the description is general enough to be comparable with other descriptions (allowing for cross-linguistic research), but specific enough to capture the uniqueness of the language (Lehmann 1999: 6). While language descriptions are extremely important tools for linguists, they have a more limited range of uses for people outside the linguistic community. This is not to say that language descriptions do not, for example, benefit the community of the language. The fact that the language has been described, that books such as grammars and dictionaries "prove" that the language is viable and worthy of respect (Payne 1997), has actually helped lessen the stigmatization of endangered languages and helped raise awareness for the value of them, which in turn has led to a higher sense of pride of the communities for the languages. The combined effect of descriptions and documentations may thus lead to a higher recognition, even on a political level, of the language. Reference materials, such

as educational material, produced in combination with descriptive materials, might lead to a higher awareness, and might even lead to the language becoming recognized enough to be taught in schools. This might lead to the language getting a higher status, which in turn might lead to fewer people switching away from it. For a very accessible guide on how to describe a language, see, for example, Payne (1997). Very helpful tips and suggestions on the various methodological issues involved in writing grammars are the collected papers in Payne & Weber (2007).

3.2 Sampling

As must have become abundantly clear by now, typological surveys are dependent on data from different languages, often from a large number of languages. To include all human languages in an investigation is simply not possible, as discussed above. Not only do we not have access to all human languages, since we have data for only very few of those languages that have already gone extinct, and no data at all for those that are yet to emerge, but we also have limited access to the existing languages of the world. While new descriptions of previously undescribed languages are constantly being done, the majority of the world's languages still remain either poorly described or not described at all. Conducting fieldwork on the thousands of languages that remain poorly or not at all documented in order to include them in one's investigation is not humanly possible. Therefore one must choose a sample of languages and infer the patterns from that sample. It is thus important to keep in mind at all times that statements about cross-linguistic patterns, tendencies and universals are always based on a sample of languages. In order to make a sample as representative as possible, a number of factors need to be considered. This section will outline the basic principles for different kinds of samples. It is by necessity a brief outline; for a discussion on language sampling, see Bakker (2010), which this section relies a great deal on. A detailed discussion of sampling methods and the various issues related to them can also be found in Song (2001: 17ff).

3.2.1 Types of samples

The type of sample to choose largely depends on the type of investigation to be conducted. In order to check for statistical tendencies and correlations of various features, that is, whether languages tend to prefer a certain word order or a certain kind of morphological strategy, and so on, we need a **probability sample**. Here we have set the **variables** (also called **features**, **characters**, or **parameters**, among other terms) beforehand and map the sample according to presence or absence of those variables. For example, we may want to check patterns for reduplication. We may then choose a set

of variables, such as (*i*) the language does not have reduplication; (*ii*) the language has partial reduplication only; (*iii*) the language has full reduplication only; and (*iv*) the language has both partial and full reduplication. We then proceed to code each language of the sample according to those variables, choosing only one variable per language. For this kind of sample it is especially important to have genealogical and areal balance, since the strategy a language has is likely to be inherited, but may also be prone to diffusion (to spread via language contact). A genetically or areally biased sample would then give biased numbers and patterns. Types of bias will be discussed below.

A **variety sample** is "mainly used for explorative research: when little is known about the form or construction under investigation it is important that the sample offers a maximum degree of variation of the linguistic parameters [i.e. variables] involved" (Rijkhoff & Bakker 1998: 265). In a sense this can be likened to a kind of snapshot of the diversity and distribution of a feature. Therefore this kind of sample must above all seek to include all possible (known) variables. This is generally done in stages: at first the linguist may not know what kinds of variation there might be for the linguistic feature under investigation, so s/he sets up a genetically and areally diverse sample. If it turns out that there are many unique cases for the feature, the sample may be enlargened to try and capture undetected values. Here languages may have the same origin, as long as they have different values for the feature in question; this may result in genealogically or areally biased samples (cf. Bickel 2008). What is crucial is to try and capture all (or at least as many as possible) values for the feature that is being investigated.

A **convenience sample** is a sample based on what kind of data one has access to. While a typologist may set up as fancy a sample as s/he likes, in the end access to data will play a big role in what goes in. Obviously, if there is no information available for the desired language, it simply cannot be included in the sample. However, while we are not able to select data for which there are no descriptions (and hence are to various degrees forced to construct convenience samples), we can still try to make these convenience samples as genealogically and areally balanced as possible.

3.2.2 Types of bias

In order for a sample to be as representative as possible, we wish to avoid that it is biased, since a biased sample is likely to affect the result of the survey in various ways. For language samples there are various factors that can make a sample biased, the most significant ones being genetic, areal, typological, cultural, and, very frustratingly, **bibliographical** bias. As has been repeatedly mentioned, adequate linguistic descriptions are lacking for the majority of the languages of the world. The descriptions that we do have tend to be of well-known, accessible languages, which means that small or remotely located languages, very often isolates or languages of unknown affiliations,

are (bibliographically) biased towards exclusion from the samples. For instance, as mentioned in Chapter 2, until Derbyshire (1977) published his description of the Hixkaryana word order, object initial languages were not to be found in any surveys on types of word order. Likewise, any survey checking for how many stop contrasts languages could have in the coronal area would have assumed that languages could only have a maximum of three contrasts, until Heath's (1984) grammar of Nunggubuyu (Australian (Nunggubuyu): Australia) was published, showing that Nunggubuyu has four phonemic coronal stops: dental, alveolar, palato-alveolar and retroflex.

A **genetic** (or genealogical) bias means that some language families are overrepresented while others are underrepresented in the sample. This is in many cases an effect of bibliographical bias, since a researcher will include the material that is available. But if material is primarily available from a small selection of language families, the sample will be biased to that effect. For example, there is a great deal of material available for Indo-European languages, especially European ones, but less available for Khoisan or Australian languages.

Many features of a language are inherited. If a sample is biased towards one family over others, a feature might look more or less common than it actually is, simply because of how it appears in the dominating family. Tone, for example, is not a common feature in Indo-European languages, but it is quite common in Niger-Congo languages. If a sample has a higher proportion of Indo-European languages than other families, the pattern that is likely to emerge is that tone seems less common cross-linguistically than it actually is (which, in fact, is the case with Maddieson's 2011f survey, as he himself points out).

An **areal** bias means that languages from the same linguistic area are overrepresented, which again may skew the resulting pattern one way or another. Linguistic areas (*Sprachbunde*) are areas where languages have been in sustained contact and have influenced each other so that they have specific features not found in the languages outside the area (see further 13.2). For example, the languages of the Balkan area, which belong to different genera of Indo-European, have postposed articles as opposed to the neighbouring languages outside the linguistic area, and as opposed to other languages of the same genera. They also have a specific type of analytic future tense, again as opposed to the languages outside the Balkan area, or the other languages of their genera. It would be premature to base any kind of hypothesis as to the correlation between these two features before checking if other languages in the world also exhibit similar kinds of correlations, because the reason that the two features emerged among the languages in the Balkan area is probably due to the shared history and contact between the languages, not because these two phenomena tend to go hand-in-hand cross-linguistically.

A **typological** bias means that one linguistic type is over- or underrepresented in a sample. For example, if we want to investigate an implication of some kind ("if X, then Y"), then we need to make sure that the various known types of X and Y are represented. Thus, if we want to check if there is any correlation between adposition and verb-object word order, we need to include languages of all types, i.e. those with prepositions, those with postpositions, and those with inpositions and check each of them against languages with verb-object word order as well as check each of them with languages with object-verb word order. If we have an overrepresentation of languages with, for example, prepositions, we are likely to get a skewed pattern.

With **cultural** bias we have an over- or underrepresentation of the different cultures of the world in the sample. It seems reasonable to assume that there is "a relation between certain aspects of the grammar of a language on the one hand and beliefs and practices of its speakers on the other hand" (Bakker 2010: 108). For example, in a study on number marking, Lucy (1992) found that speakers of American English and speakers of Yucatec (Mayan (Mayan): Mexico) treat nouns differently: the English speakers make a sharp distinction between mass and count nouns, and have obligatory number marking for count nouns, whereas the Yucatec speakers treat most nouns as mass nouns, have optional number marking but an obligatory numeral classifier system (based primarily on the material of the object). In extra-linguistic tasks, such as looking at pictures and then later describing them from memory, the speakers of the two groups paid attention to and remembered different things. The English speakers were more likely than the Yucatec speakers to remember the number of objects on the pictures, especially with respect to objects that constitute count nouns. Also, while the English speakers were more likely to pay attention to and remember the shape of the different objects, the Yucatec were more likely to pay attention to and remember the material composition of the different objects. Similarly, when asked to sort pictures of objects, the English speakers tended to sort objects by shape, while the Yucatec speakers tended to sort objects by material composition. Thus the way the culture interprets objects, and the way the language of the culture marks for number, seems to correlate.[26]

It is not likely that we will ever be able to create a completely balanced sample. For one thing, bibliographical bias will continue to affect the choice of languages for a sample. However, most of us try to control for areal and, above all, genealogical bias, the latter being especially relevant because "if languages are closely related genetically, they are likely to have inherited common linguistic types from their ancestor language,

26. Whether the number marking system derives from the cultural outlook (i.e. how one views and categorizes objects) or whether the cultural interpretation of objects derives from the linguistic structure, is probably impossible to establish. In any case the culture and language of a community go hand-in-hand and are expressions of each other.

to be spoken in the same area and by people sharing the same culture" (Cristofaro 2005: 91).

Ideally, of course, statistically valid statements should be based on **randomly** sampled data. But because of the nature of linguistic data, this is not practical, since this might lead to areally skewed patterns, as we might end up choosing languages that are in intense contact and might therefore share features due to diffusion rather than inheritance or anything else:

> for example, when random sampling happens to pick Romansch [(Indo-European: Switzerland)] as the sole representative of Romance and High Alemannic [(Indo-European: Switzerland)] as the sole representative of Germanic, this will overestimate areal effects in Europe because the two languages are under much more intense contact with each other than, say Portuguese and Swedish. In response to this one might chose to admit several languages from each stratum [i.e. genus] in the hope of reducing such effects. However, this option is severely limited because about a third of the proven stocks [i.e. families] in the world are isolates. Since strata [i.e. genera] need to contain the same number of languages, the inclusion of isolates implies that only one datapoint can be admitted for each stratum, even for non-isolates like Romance and Germanic. (Bickel 2008: 222)

The above two sections serve to highlight the various problems and considerations involved in constructing samples and trying to avoid bias. As mentioned above, a sample completely free of bias is unlikely to come by. It should be noted that I will be making extensive use of existing databases and surveys and will be quoting the patterns found therein wholesale, without controlling for bias. In other words, I leave the responsibility of the sample compilation to the author(s) quoted. The reader should therefore be aware of the potential bias in samples quoted (for a discussion on WALS and APiCS in particular, see below). This should in no way be taken to imply that a biased sample is without any value. On the contrary: it still brings us forward in our understanding of the complexity of language.

On a final note, statistics may seem far removed from typology, but is actually pretty essential, since what we are dealing with is sets of data, samples aimed at representing the whole (the so-called 'population', i.e. the entire collection of what we are investigating, in our case the languages of the world), and drawing conclusions from these sampled data. For an introduction to statistics for linguists, see, for example, Gries (2009).

3.3 Databases

For the last number of years databases have increasingly become publicly available, a practice which is highly beneficial for both compilers and the linguistic community (as well as the general public) and is a trend that should be encouraged. Publicly available databases are a great benefit for research, making scores of data accessible. They also allow the compilers to be recognized for their painstaking work. Another great advantage of making databases available online is that they can continuously be updated; not only can errors that have been detected be corrected, but as new material becomes available and as new insights emerge, data can be added or modified according to the latest state of affairs.

Any database will involve compromises, which means that databases may differ radically from each other, both in selection of languages and in the approach to the entries. There are databases with a vast amount of languages, but where the data provided for each language is quite restricted. There are databases providing very elaborate information for each language, but the number of languages is then smaller. There are databases which look only at one specific linguistic domain, while other databases code a host of features and information about the language. The following will discuss the respective merits and shortcomings of three quite different kinds of databases that are publicly available, all of which I have been or am personally part of and thus have direct experience with not only as a user but also as a contributor. There are many more very informative and valuable databases either already publicly available or projected to be available in the near future (see Appendix 1).

A milestone in terms of large-scale databases is the *World Atlas of Language Structures* (WALS), which started in 1999 as a collaborative effort between experts of different linguistic features. What WALS did was essentially to pull together a number of databases into one single unit, covering a great part of the abstract linguistic system of a language. The chapters range from phonological features, over morphology, grammar, and syntax, to lexical features and provide the first world-wide collected mapping of language systems. Another groundbreaking aspect of WALS is its inclusion of two chapters on sign languages. For WALS each linguistic feature was dealt with separately, i.e. authors used their definitions and expertise to map a specific feature and each chapter basically consisted of one feature. To make the maps cross-compatible, a sample of 100 languages that each chapter had to include was set up by the editors of the project. An extended sample of a further 100 languages (making the total 200) was encouraged to be included if possible. The instructions for the authors were basically to map their features for a set core sample of languages. The sample was chosen with the intention to maximize areal and genealogical distribution of the languages within the confines of available descriptions. Thus a minimal core sample of 100 languages allows for

investigations of correlations between features. Very often many more languages overlap between the chapters. Furthermore, the atlas provides metadata for each language included, specifically the location of the language and its genealogical classification.

Because authors were responsible for individual features, their chapters may contain a large amount of languages, though these may not necessarily overlap with other chapters. As a result, even though the total amount of languages in WALS is very large, with 2559 for the printed version and 2678 for the current online version, the range of overlap between chapters can vary considerably. This means that despite the high number of total languages, there are instances when combinations of features only yield the obligatory 100 sample. In other words, while one can find information on all the 142 (144 in the current online version) features for some languages, such as English, Turkish (Altaic (Turkic): Turkey) or Lavukaleve (Solomons East Papuan (Lavukaleve): Solomon Islands), there are other languages that only occur in one or a few chapters. As mentioned above, while the core sample was as balanced as possible at the time (1999), it was by necessity a convenience sample and therefore has a certain amount of bias towards Eurasian families (for a discussion on the balance of the WALS sample, see Hammarström 2009). Furthermore, it almost completely ignored pidgin, creole and mixed languages.

In contrast to WALS, in the *Atlas of Pidgin and Creole Language Structures* (APiCS), the first large-scale typological project for pidgin and creole languages,[27] the authors were responsible for a language while the features that were being investigated were set by the editors. Here, again, the features cover a great part of the abstract linguistic system, ranging from phonological analysis, via morphology, grammar, syntax to the lexicon. This project basically pulls together experts on different pidgin, creole and mixed languages. Because the features are predefined, and authors are responsible for specific languages, the cross-compatibility between languages is absolute. That is, the kind of information that can be found for one language can essentially be found for every language in the database. The instructions for the authors were to fill out a detailed questionnaire of features for the language of their expertise. Each language is also described in a survey chapter, containing a summary of the socio-historical background, as well as a broad structural outline of the language. This amount of details provided for each language necessarily reduces the amount of languages that can be included in the database – experts have to be found that are willing to invest the time and effort in the project – and the APiCS includes only 76 languages, i.e. a fraction of the number of languages in WALS.

---

27. Other surveys across pidgin and creole languages have of course been made, but not of the same scale. A kind of a predecessor to APiCS, containing many features that eventually made it into the APiCS questionnaire, is the volume edited by Holm & Patrick (2007), which covers various syntactic features of 18 creole languages.

The APiCS, however, includes only pidgin, creole and mixed languages, that is, selected languages that may or may not be of a specific typological sort (see above), on the assumption that there is something special about these kinds of languages. Whether they actually are special or not is difficult to ascertain if they are not compared to languages that are not considered special. While many of the APiCS features actually do overlap partially or even fully with features in WALS, thus allowing for a comparison between contact languages and 'non-contact languages', this is not true for all features. A complete cross-comparison between APiCS and WALS is therefore not possible. Furthermore, the sample of the APiCS is biased towards English-lexified contact languages (27 of 76, or more than a third of the entire sample).

An entirely different database altogether is the *Automated Similarity Judgement Program* (ASJP), which aims to provide an objective classification of the world's languages by means of lexicostatistical analysis. Lexicostatistics[28] is a technique used to quantitatively compare the rates of change within a set of words in different languages in order to try to establish in how far they are related and, if they are, when they separated from each other. The ASJP computerizes this comparison between sets of words, using a fixed algorithm. The method is therefore objective and consistent across the board, and makes the classifications arrived at more objective and consistent than the existing ones, which are all based on different experts' knowledge, and, by extension, their various hypotheses. The task for each contributor in ASJP is to enter a set of 40 lexical items for as many languages as possible. Some macro-data is included for each language, such as genealogical affiliation, location and number of speakers. Since the dataset is so small, it is possible for contributors to submit a large amount of languages. The database thus at the time of writing contains a staggering 5751 languages, dialects and proto-languages (reconstructed languages assumed to be the nearest equivalents to ancestor languages that we can arrive at) and the number is constantly increasing. This is to date one of the largest cross-linguistic databases available.

Since we are dealing with a computerized comparison, the words have to be transcribed in a machine readable format. This is done by approximating the original transcriptions of the words in the sources according to a set of transcription rules. That means that the data is a reduced form of the original. It may thus not be of much use outside the project. And since the transcription is an approximation of the original, it is not possible to simply convert it back to a more detailed format in order to make it more accessible to others. We thus have a database with a vast amount of languages, each only containing a small dataset which is of limited use outside the project.

---

28. Sometimes also called glottochronology. For those who make a distinction between them, glottochronology is specifically concerned with establishing a date for the split between languages into daughter languages, while lexicostatistics is "the statistical manipulation of lexical material for historical inferences (not necessarily associated with dates)" (Campbell 2004: 201).

3.4 A note on methodological issues with sign languages

The issues discussed in this chapter all hold for sign languages too. However, due to the lack of descriptions and availability of data, cross-linguistic surveys of sign languages by necessity differ in their approaches to methodological issues for cross-linguistic surveys of spoken languages.

Just as linguists making cross-linguistic surveys of spoken languages have to rely on various methods of accessing data, such as through language descriptions, elicitation and questionnaires, as well as texts, so do linguists making cross-linguistic surveys of sign languages. Because sign language descriptions are still sorely lacking, much of the information that goes into databases relies on expert knowledge, questionnaires or other forms of elicitation. To give an example, the data for Zeshan's (2011a and 2011b) chapters in WALS rely to a great extent on personal communication with experts.

Because of the lack of data for sign languages, sampling is not really relevant for cross-linguistic surveys. Rather, as yet any and all available information gets included in the surveys (Zeshan 2011c). In other words, due to the status of available information on sign languages, typological surveys of them are by force convenience samples. This of course means that all available surveys at this stage are biased; most available descriptions are of European sign languages.

To document and describe sign languages is therefore a matter of urgent need. The mechanisms and considerations involved are largely the same as with documenting and describing spoken languages: fieldwork needs to be conducted, consultants need to be found, forms and methods for documentation need to be considered, descriptions based on these documentations need to be done (preferably as free of predefined hypotheses and assumptions as possible). The major difference in procedure between spoken and signed languages is that signed languages are three-dimensional and visual/gestural languages, while spoken languages are two-dimensional and audio/oral languages. For obvious reasons, then, the fieldwork equipment needs to be visually based (i.e. video cameras) and not audio based. Descriptions of sign languages are best done in a three-dimensional medium, which is quite straightforward nowadays with DVDs and the internet readily accessible to a large number of users, although most currently available descriptions are in written form.

There are a number of endangered sign languages. The current (16th) edition of the *Ethnologue* lists Maritime Sign Language (MSL: Canada), Providencia Sign Language (PSL: Providencia Island), Hawai'i Pidgin Sign Language (HPSL: Hawaiian Islands)[29],

29. This is not a 'pidgin sign language' in the sense of International Sign, which is a contact language "that has developed in international settings as a result of contact between deaf users of different sign languages from Europe and North America" (Zeshan 2011c), but simply a locally based sign language called *(Hawai'i) Pidgin Sign Language* by others.

Benkala Sign Language (also known as Kata Kolok, KK: Indonesia) and Rennellese Sign Language (RSL: Solomon Islands) as nearly extinct (Lewis 2009), though this clearly does not cover all known endangered sign languages. Plains-Indian Sign Language (PISL: USA), for example, is endangered by now (Zeshan 2011c). AdaSL, an indigenous sign language of Ghana, is rapidly losing native signers, since, due to the schooling system, signers are shifting to Ghanaian Sign Language (GSL; Nyst 2007). Similarly, signers of the indigenous Ban Khor Sign Language (BKSL) in Thailand are shifting to TSL, while other indigenous sign languages of Thailand, such as Old Bangkok Sign Language (OBSL) and Old Chiangmai Sign Language (OCSL) are near extinction, sporadically remembered by only a few signers over 50 years of age, but no longer used in daily conversation (Nonaka 2004: 743ff).

3.5 Summary

Any linguistic research depends on language data. Typological surveys often involve a high number of languages. Data is typically obtained from language descriptions, through native speakers or specialists by way of elicitations or questionnaires, or through texts spoken or written by native speakers.

Languages are disappearing at an alarming rate, with roughly half of the languages of the world projected to have gone extinct by the end of this century. Language documentation is an urgent task. Any kind of language documentation involves fieldwork in some form. Documentation and description of a language means collecting data and providing an interlinearized glossing and translation of it. Describing a language means using the documented data to provide an analysis of the language.

Due to the impossibility of including all languages in a typological survey, samples have to be made. Probability samples have a set number of values that are mapped across languages. Variety samples seek to capture all possible variables of a feature. Convenience samples are samples based on access to data. Completely random samples are rarely practical for linguistic surveys due to the risk of inflationary areal effects.

Samples can be biased in various ways. Bibliographically biased samples are those where imbalance is due to lack of information. Genealogically biased samples are those where language families or genera are either over- or underrepresented. Areally biased samples show over- or underrepresentation of languages from specific linguistic areas. Typologically biased samples show an over- or underrepresentation of certain linguistic types. Culturally biased samples show an over- or underrepresentation of various cultures in the sample.

Lack of data for sign languages forces all surveys to be convenience samples. This by necessity makes the surveys biased. There is an urgent need for documentation and description of sign languages. Documentation and description of sign languages essentially involves the same kinds

of considerations as with spoken languages, with the exception that the fieldwork equipment has to be able to capture the gestural/visual nature of sign languages. A number of known sign languages are endangered at this stage.

3.6 Keywords

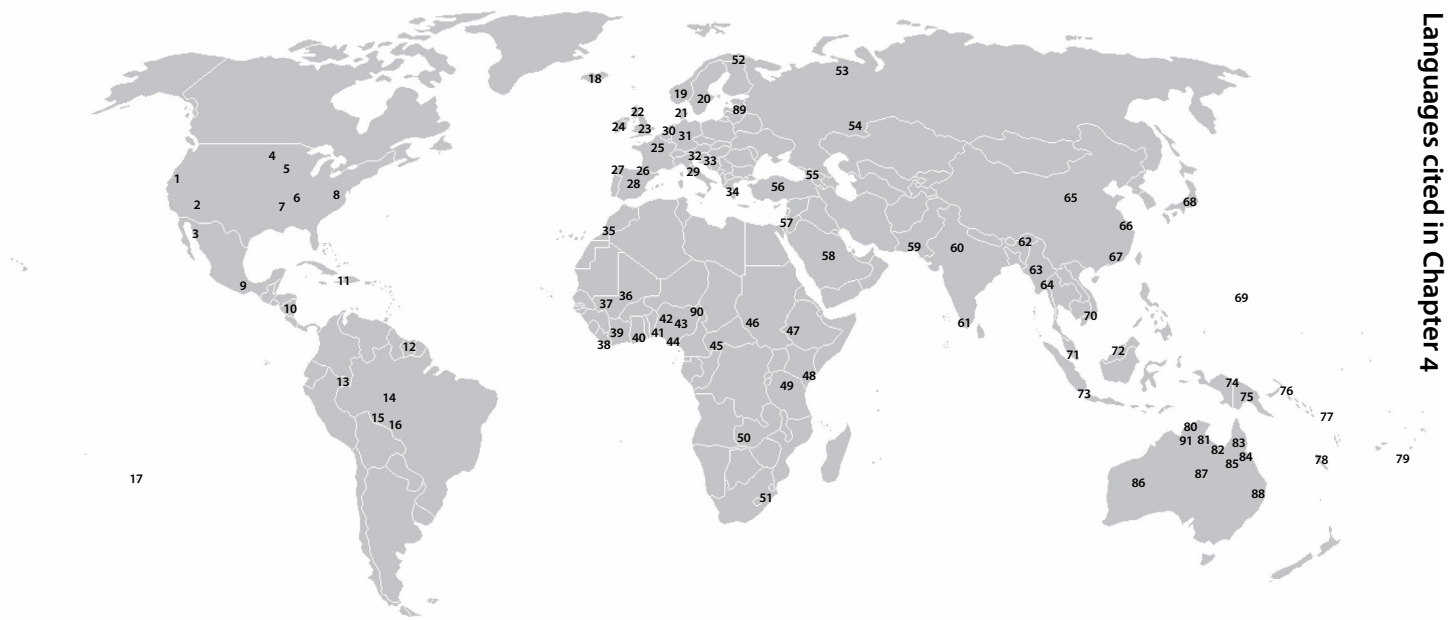
data source
database
endangered language
fieldwork
language description
language documentation
moribund language
sampling
sample bias
stable language

3.7 Exercises

1. Describe the four main types of data sources for typological surveys and state whether (and if so how) they differ between spoken and signed languages.
2. What are endangered and moribund languages?
3. What is the difference between a probability, variety and convenience sample?
4. What kinds of bias are the most significant for typological surveys and why?
5. Is the following statement true or false? Motivate your answer.

   A typological survey can never cover all human languages.

Languages cited in Chapter 4



1 Yurok
2 Hopi
3 Seri
4 Dakota
5 Winnebago
6 Quapaw
7 Wichita
8 American Sign Language
9 Chinantec, Quiotepe
Zapotec, Tilquiapan
10 Rama
11 Haitian Creole
12 Sranan
13 Ticuna
14 Pirahã
15 Wari'
16 Aikaná
17 Rapanui
18 Icelandic
19 Norwegian
20 Swedish
21 Danish
22 Gaelic, Scots
23 English
24 Irish
25 French
26 Basque, Goizueta
27 Galician
28 Spanish
29 Italian
30 Dutch Sign Language
31 German
German Sign Language
32 Slovene
33 Croatian
34 Greek
35 Tashlhiyt
36 Koyra Chiini
37 Bambara
38 Klao
39 Bété
40 Adamorobe Sign Language
41 Yoruba
42 Bauchi Guda
43 Fyem
44 Ejagham
45 Sango
46 Dinka, Lanyjang
47 Gimira'
48 Dahalo
49 Hadza
Sandawe
51 Zulu
52 Saami, Inari
53 Khanty
54 Bashkir
55 Georgian
56 Turkish
57 Al-Sayyid Bedouin Sign Language
Hebrew
Israeli Sign Language
58 Arabic, Standard
59 Urdu
60 Hindi
61 Malayalam
62 Nishi
63 Burmese
64 Kayah Li, Eastern
65 Chinese, Mandarin
66 Wu, Changzhou
67 Hakka
68 Japanese
69 Chamorro
70 Vietnamese
71 Singlish
72 Berawan, Central
73 Batavia Creole
74 Dumo
75 Hua
Yimas
76 Rotokas
77 Natügu
78 Iaai
79 Tongan
80 Malakmalak
81 Yanuyuwa'
82 Kayardild
Lardil
83 Kunjen
84 Dyirbal
Yidiny
85 Mbabaram
86 Mantjiltjara
87 Arrernte, Mparntwe
88 Bandjalang
89 Latvian
90 Margi
91 Wardaman

Chapter 4

Phonology

Natural languages make use of contrastive elements that combine to form meaningful units such as words. Spoken languages make systematic use of sounds and arrangements of these sounds to form words and utterances. This chapter gives an overview of the various sounds found in the world's languages and how they tend to pattern. I will first give a sketch of different kinds of phonemes (4.1.1), then bring up how they combine sequentially to form larger units, syllables, and then move on to mention the role of prosody (or suprasegments) (4.1.2). Section 4.2 gives an overview of cross-linguistic patterns, starting with segment inventories and some common as well as less common phonemes (4.2.1), after which patterns of syllable structures will be brought up (4.2.2). Section 4.2.3 will mention some patterns of prosody, specifically length, tone and stress. Section 4.3 gives an overview on sign language phonology.

4.1 Phonemes, syllables and prosody

The following is meant as a quick overview of the very basics of phonology and is by no means exhaustive. For very accessible introductions to phonetics and phonology, see e.g. Ladefoged & Johnson (2010), Ladefoged (2005) and Spencer (1996).

The study of sounds involved in spoken language falls under **phonetics** and **phonology**. Very broadly, the former deals with the physical characteristics of speech, and links linguistics with other disciplines, such as physics (the acoustic dimensions of speech), anatomy (the physiological aspects of speech production), and psychology and neurology (the auditory aspects, primarily of speech perception). The latter, phonology, deals with the abstract linguistic patterning of the units produced in speech. Basically, that means that phonology studies how speech sounds are used by speakers to mark linguistic contrasts, i.e. how sound is used to convey one meaning and not the other. For instance, in phonetic terms, the initial consonant in *pie* [pʰai] is not exactly the same as the second consonant in *spy* [spai], but in phonological terms they are both variants (**allophones**) of the **phoneme** /p/. This phoneme then contrasts with other phonemes, such as /t/ in *tie* [tʰai] and *sty* [stai]. (To be able to distinguish

the two from each other, phonetic transcriptions are given in square brackets while phonological transcriptions are given between slashes.)[30]

It should be borne in mind that no two speakers produce exactly the same kinds of sounds when they speak. Likewise languages that are said to have the same kinds of phonemes do not produce exactly the same kinds of sounds in phonetic terms. For example, the five Spanish vowels, /a, e, i, o, u/, are, in absolute terms, not exactly the same as the five Japanese vowels which are also transcribed /a, e, i, o, u/. However, in both languages, the way the vowels contrast with each other on a vowel chart is roughly similar and so the five vowels of each system are viewed as having phonological properties that are similar enough to warrant transcribing them with the same symbols. Throughout this chapter it should be kept in mind that we are dealing with contrastive (phonological) segments and not absolute (phonetic) qualities.

Phonemes are the smallest meaning distinguishing units of a spoken language. Languages tend to have **minimal pairs**, where a meaning is changed by swapping only one single phoneme in the same position, such as ***p**ot* versus ***c**ot*. Groups of words like ***h**at*, ***b**at*, ***c**at*, ***m**at*, or *b**i**t*, *b**u**t*, *b**a**t*, form **minimal sets**. A very important point to bear in mind is that what is contrastive in one language does not necessarily have to be contrastive in another language. For instance, while [pʰ] and [p] are not contrastive in English, as we saw above, but are allophones of the same phoneme /p/, this is not the case in Hindi (Indo-European (Indic): India), where we find such minimal pairs as /pal/ 'nature' versus /pʰal/ 'knife blade' (Ohala 1994: 35).

Spoken languages make use of **consonants** and **vowels**. Vowels are formed when air passes freely from the lungs through the mouth. Slight movements of the tongue change the shape of the oral cavity which leads to different vowel qualities. Consonants are produced by creating obstacles for the air in the vocal tract in various ways. These obstacles can be created in different places (**places of articulation**) and in different ways (**manners of articulation**).

4.1.1 Segments

The standard way of representing the sounds of spoken languages is with the International Phonetic Alphabet (IPA). The IPA chart regularly gets revised, adding symbols for sounds previously not thought to occur in languages (a case in point being the labiodental flap (see below), which was only added in the 2005 version of the

30. The superscript ʰ indicates that the consonant is **aspirated**, i.e. that a short burst of air accompanies it. With post-aspirated consonants the burst of air comes immediately after the consonant has been articulated. Languages may also have pre-aspirated consonants, in which case the burst of air immediately precedes the articulation of the consonant. Icelandic (Indo-European (Germanic): Iceland), for example, has preaspirated consonants, such as /saʰka/ 'sinkstone' or /saʰkna/ 'to regret' (Ladefoged & Maddieson 1996: 71).

chart), and taking out symbols as analyses change (for instance the voiceless implosives, which are no longer in the chart).

The current chart is given in Figure 4.1. I will use it as a means for giving the basics needed in order to understand Section 4.2. As I will be making use of the IPA notational system, the reader should refer to Figure 4.1 as a guide to the symbols. Bear in mind that this is a very simplified overview; I will not explain every item listed in the chart, but merely those necessary to follow the subsequent discussion. For a thorough overview of the sounds in the world's languages, see Ladefoged & Maddieson (1996). The remainder of this chapter relies heavily on the phonological data of the 451 languages in *The UCLA Phonological Segment Inventory Database* (UPSID). I have used the web interface developed by Henning Reeth (http://web.phonetik.uni-frankfurt.de/upsid.html, accessed in March 2010).

4.1.1.1 *Consonants*

The obstacles created to produce consonants can vary in degree. **Obstruents** are produced by a high degree of constriction of the airflow whereas **sonorants** allow a relatively free airflow. If the air is completely closed off, **plosives** (or **stops**) are produced. All languages have plosives. If the two main **articulators** (the main two parts of the speech apparatus forming the sound) are placed very close together so that they obstruct the flow of air without closing it off completely, we get **fricatives**. While fricatives are very common, about 8.5% of the languages in Maddieson's (2011a) sample lack them, most of them Australian, such as Wardaman (Yangmanic: Australia; Merlan 1994: 11). **Affricates** are essentially consonants that start out as stops and end as fricatives. About 66.5% of the languages in UPSID have them. Stops, fricatives and affricates are obstruents.

Sonorant consonants include **nasals** and **approximants**. Nasals are produced by allowing air to flow out of the nasal passage during the articulation. Nasals are very common indeed, with only 10 languages in Maddieson's (2011a) sample, such as Rotokas, lacking them. Approximants are formed by causing such a low degree of obstruction to the air that friction is not produced. They are also very common, present in 96.2% of the languages in the UPSID. Approximants are sometimes divided into **liquids** and **glides**. Liquids are basically "r-sounds" (or **rhotics**) and "l-sounds" (**laterals**). Laterals are formed by closing off a band in the middle section of the mouth and letting the air flow on one or both sides of the tongue. Rhotics can be either **taps** or **flaps**, which are formed by flicking the tip of the tongue against some point of the roof of the mouth, or **trills**, where the tongue is rolled against the place of articulation. Glides (or **semi-vowels**) are such sounds as *w* and *j*.

If the vocal folds vibrate during the articulation, a **voiced** consonant is produced, if they don't, the consonant is **voiceless**. All languages in the UPSID contain some kind of

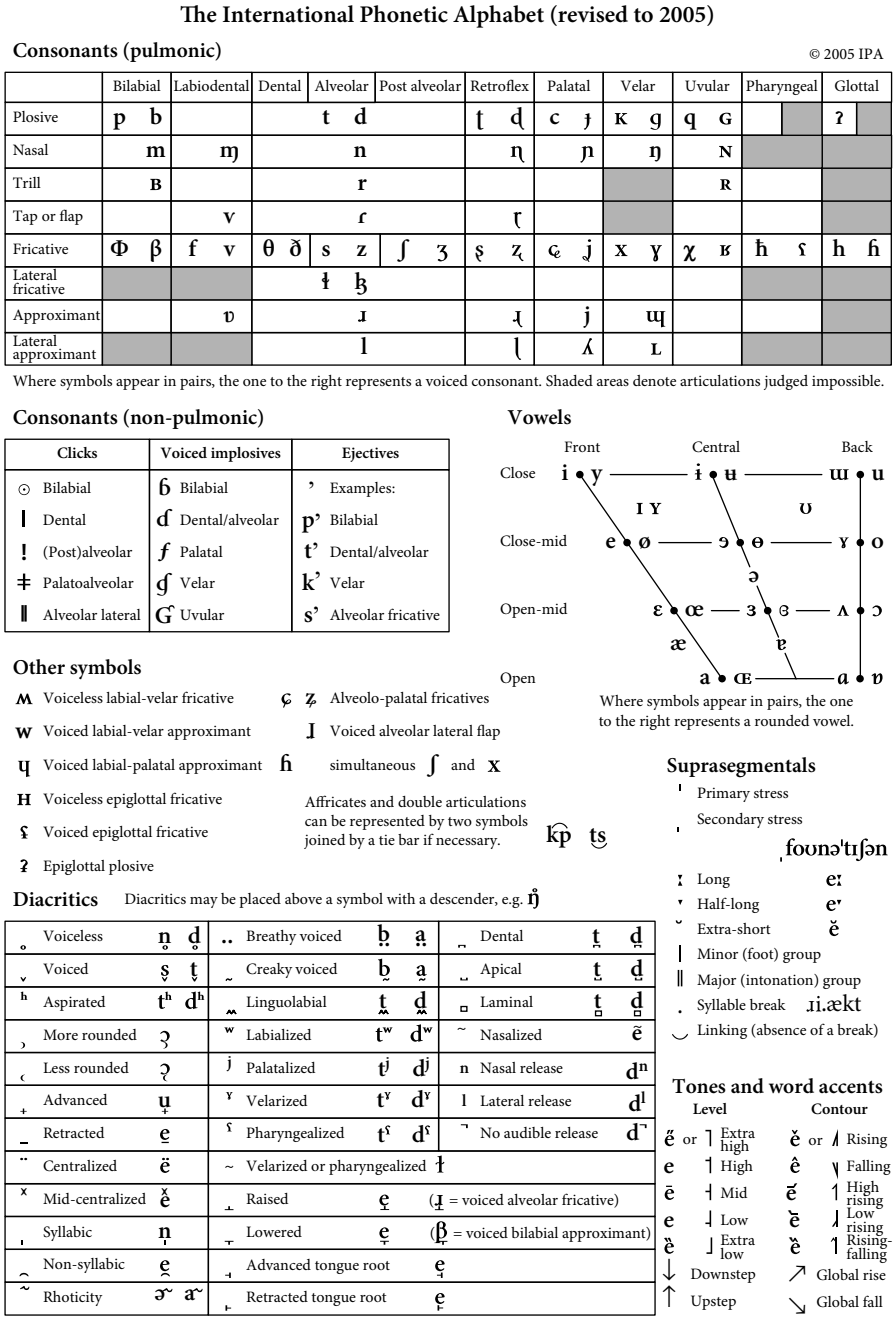


Figure 4.1 The full IPA chart (http://www.langsci.ucl.ac.uk/ipa/ipachart.html accessed: 9 April 2010).

voiced consonant, but four Australian Pama-Nyungan languages (Bandjalang, Dyirbal, Mbabaram and Yidiny) in UPSID lack voiceless consonants.[31]

Starting from the front of the mouth and working our way backwards, **labial** consonants are produced with the lips. There are two types of labials, those involving both lips, **bilabials**, and those involving the upper jaw front teeth and the lower lip, **labiodentals**. Bilabials are extremely common: a full 563 of 567 (99.3%) languages in Maddieson's (2011a) sample have bilabials. Wichita (Caddoan (Caddoan): USA) is an example of a language that lacks bilabials (Rood 1996; note, however, that bilabials are only found in 45% of the languages in UPSID).

Consonants produced by raising the front part of the tongue (either the tip or the blade) towards the roof of the mouth are called **coronals**. These comprise the biggest group of consonants. If the tip of the tongue is used, the sound is **apical**, if the blade of the tongue is used the sound is **laminal**. There are languages where this is the only contrastive feature, such as Yanuyuwa (Australian (Pama-Nyungan): Australia) (see below). If the underside of the tip of the tongue is used, the sound is **retroflex** (or **sub-apical**). **Dentals** are produced by raising the tongue to the back of the front teeth. About a third (35%) of the languages in UPSID have them. **Alveolars** are produced by raising the tongue to the hard dome-like structure immediately behind and above the front teeth, the alveolar ridge. Alveolars are common sounds, found in 63.9% of the languages of UPSID.[32] **Post-alveolars** are formed by raising the tongue to the back of the alveolar ridge (but not as far back as the hard palate). These are also common sounds (found in about two thirds or 64.1% of the languages in UPSID). Yanuyuwa is an example of a language where an apical post-alveolar contrasts with a laminal post-alveolar as in /wuḍulu/ 'in the stomach' and /wuḏulu/ 'into the grass' (Ladefoged & Maddieson 1996: 35). The crucial feature of retroflex consonants is that they are formed by using the underside of the tip of the tongue to touch an area roughly between the post-alveolar and palatal section of the mouth. About a fifth (20.2%) of the languages in UPSID has them and they are especially common in, but not limited to, South Asia.

Consonants produced with the body of the tongue are called **dorsals**. **Palatals** (found in 89.8% of the languages of UPSID) are formed by raising the body of the

---

31. Sonorants may also be voiceless, such as the nasals of Burmese (Sino-Tibetan (Burmese-Lolo): Myanmar) (cf. for instance the minimal pair /nă/ 'pain' versus /n̥ă/ 'nose'; Bhaskararao & Ladefoged 1991: 80). Icelandic nasals occurring before voiceless stops are also voiceless and have been argued to function as phonemes (see Jessen & Pétursson 1998). Even vowels may be voiceless, which means they are then in effect whispered, as in Japanese /ki̥ʃi/ 'shore' (Ladefoged & Maddieson 1996: 315).

32. It should be noted that a very large number of languages in UPSID (207 or 45.90%) have been coded dental/alveolar, meaning that it is not specified whether the sound is dental or alveolar. The figures mentioned above for dentals and alveolars respectively are therefore not entirely representative for the languages in the database.

tongue to the hard section of the roof of the mouth, the palate. Further back is the soft section of the roof of the mouth, the soft palate or velum, where **velar** consonants (found in 99.6% of the languages in UPSID) are formed. Only two languages in the database do not have any velars: Dumo (Skou (Western Skou): Papua New Guinea) and Klao (Niger-Congo (Kru): Liberia; which, however, does have the labio-velars /k͡b/ and /g͡b/). Consonants produced with the soft little piece of flesh dangling at the very back of the mouth, the uvula, are called **uvulars**. Less than a fifth (17.5%) of the languages in Maddieson's (2011g) database has uvular consonants of some kind.

Radicals are produced with the root of the tongue. Sounds produced with the root of the tongue moving towards the pharynx and the very back of the mouth are called **pharyngeals**. They are quite rare, found in only 21 languages (3.7%) in Maddieson's (2011d) sample, mainly in the Middle East and the Caucasus.

The larynx is basically the area around the vocal folds. Laryngeal consonants are produced with the 'glottis' (the space between the vocal folds) and are called **glottals**. They are found in 74.5% of the languages in UPSID.

Consonants that do not involve pushing air out of the lungs are called **non-pulmonic** (as opposed to the **pulmonic** consonants above). There are three kinds of non-pulmonic consonants, **clicks**, **implosives** and **ejectives**. Clicks and implosives involve sucking air into the mouth. Clicks are formed by the tongue making both a front and a back point of contact in the roof of the mouth, with the middle lowered to form a small pocket of air. When drawing the tongue down and back, the air is sucked into the mouth and the release makes a small clack (or click) like sound. These are very rare consonants, found in only 9 (or 1.6%) of the languages in Maddieson's (2011d) sample, almost all of them located in southern Africa.[33] Implosives are produced by moving the vocal folds downward which causes the air to move inward for a short instant. They occur in 11.8% of the languages in UPSID, predominantly in a belt across Africa south of the Sahara, but also in a cluster on the South East Asian mainland and in a scattering of languages in both South and North America. Ejectives are formed by closing off the glottis and raising the larynx at the time of release, pushing out a short and sharp burst of air from the larynx, which creates a kind of popping sound. These are also rather rare; 15.3% of the languages in UPSID have some kind of ejective (often two). They are predominantly found in North America, but there are also a number of South American languages, as well as Eastern and Southern African and a cluster of Caucasian languages, that have them.

---

33. Hale (1992: 38) reported clicks in Damin, an extinct auxiliary language (i.e. used instead of the ordinary language of the community, Lardil (Australian (Pama-Nyungan): Australia), by those novices who were about to be initiated into manhood) in Australia. There are also reports of clicks being used to 'spice up' nursery rhymes in different Chinese dialects. See further Nathan (2001).

Coarticulated consonants are sounds produced at two places of articulation at the same time. Those coarticulated consonants that are produced at two different places at the same time but in the same manner are called **doubly articulated**, as in Klao, mentioned above. Examples of this are the labiovelars /k͡p/ and /g͡b/. There are 45 (7.9%) languages in Maddieson's (2011d) sample with labiovelars, predominantly in West and Central Africa, but also in Eastern Papua New Guinea.

4.1.1.2 *Vowels*

Vowels are defined through three parameters, their **height**, their **backness** (or **frontedness**) and their **roundedness**.[34] Very simplified, **high** (or **close**) vowels are produced by raising the tongue, while **low** (or **open**) vowels are produced with the tongue lowered in the mouth. **Front** vowels are produced by moving the tongue forward, while **back** vowels are produced by retracting the tongue. **Round** vowels are formed by rounding the lips. I stress that this is an extremely simplified explanation. In actuality, the precise properties of the different vowels involve several factors, such as how the air flow through various spaces in the mouth relate to each other depending on the various movements and positions of the tongue.

Vowels can be contrastively modified in various ways. For instance with **Advanced Tongue Root** (ATR) vowels, the tongue is pushed forward. This is common in West African languages (Ladefoged & Maddieson 1996: 300ff). Holding the vocal cords only loosely together produces a **breathy voice** vowel, while holding them more tightly together produces a **creaky voice** vowel. A **rhotic** (or **rhotacized**) vowel is produced with an "r-colour" modification. Languages may also have **pharyngealized**, **strident** and **fricative** vowels (see Ladefoged & Maddieson 1996: 306ff).

A vowel produced without any change in quality during the articulation is a **monophthong**. If, however, there is a change in quality, it is a **diphthong**. The vowel in *like* /laik/ is an example of a diphthong. If there are two changes in quality during the articulation, it is a **triphthong**, as in the often heard pronunciation of *fire* /faɪə/. Languages tend to have fewer diphthongs than monophthongs (as is the case in English), but cf. Wari' (Chapacura-Wanhan (Chapacura-Wanhan): Brazil), mentioned below.

Vowels produced by letting air through the nasal passage as well as the oral cavity are called **nasalized** vowels, as opposed to **oral** vowels. About a quarter (26.2%) of the languages in Hajek's (2011) database has contrastive vowel nasalization. French is an example of such a language, as can be seen, for instance, in the pair *paix* /pɛ/ 'peace' versus *pain* /pɛ̃/ 'bread'. Languages usually have a higher number of contrastive oral vowels than nasal vowels, but Koyra Chiini (Nilo-Saharan (Songhay): Mali), for

34. I do not include length, which is a vowel **quantity** and not quality.

example, defies this pattern (Hajek 2011). Wari' is especially interesting, since it has six oral monophthongs (i.e. none of them nasal), two oral diphthongs, but eight nasal diphthongs (Everett & Kern 1997: 392ff).

4.1.2 Syllables and suprasegmentals

Segments are usually put together to form larger units, **syllables**, which in turn form words. Syllables consist of an obligatory **nucleus** (core), which is almost always a vowel, though it can be a syllabic consonant (usually a sonorant). Very often there is a **coda** consonant after the nucleus. A syllable with a coda is called a **closed** syllable, while an **open** syllable lacks a coda. The nucleus plus coda form the **rhyme** of the syllable. There may also be an **onset** consonant before the nucleus. Hakka (Sino-Tibetan (Chinese): China) is an example of a language with syllabic consonants, as in /sɻ̩/ 'teacher' (I am, for the moment, ignoring the tones of the language) where /s/ is the onset and /ɻ̩/ the nucleus (Lee & Zee 2009: 107). Languages may also allow several segments to form either an onset, a nucleus or a coda, for instance a consonant cluster at the beginning or end of the syllable, or several vowels in the nucleus (Blevins 2004: 213). There tends to be a pattern as to the internal arrangement of consonant clusters, in that the least sonorous consonants are the furthest away from the nucleus, the so-called **sonority principle**. For instance, in the word *drink* /dɹɪŋk/ the two consonants closest to the nucleus (/ɹ/) in both the onset and the coda (/ɹ/ and /ŋ/) are more sonorous than the two consonants at the ends of the syllable (/d/ and /k/).

SYLLABLES WITHOUT SONORANTS?

In Tashlhiyt (Afro-Asiatic (Berber)), a language spoken in Southern Morocco, even obstruents may be used for a syllable nucleus, as in /tʃ.tft/ 'you crushed' and /tqs.sf/ 'it shrunk'. (Ridouane 2008: 332).

Going beyond merely the segmental building blocks of spoken language, speakers also often use such **prosodic** (or **suprasegmental**) features as **duration**, **pitch**, **loudness**, and **rhythm**. The term suprasegment refers to the fact that these are contrastive features that may carry over (*supra-*) across segments. Both consonants and vowels may have contrastive durations, making them either **long** or **short**. In fact, there are languages that distinguish between three lengths (see 4.2.3.1).

Change in pitch produces various **tones**, which in many languages are as contrastive as segments. It is a **paradigmatic** feature, distinguishing units from each other. That means that the units typically substitute each other, in other words, either unit A is used, or unit B, or unit C – i.e. each form is unique. The tone may either be a flat pitch at a certain level, a **level** tone, or a **contour** tone, where the pitch moves

from one level to another (either rises or falls) over the **Tone Bearing Unit** (TBU), usually the syllable. There are also so-called **floating tones**, which are tones "associated with no syllable" (Yip 2002: 76), but which carry morphological information. This is especially common in Bantu languages. An example of a floating tone is the Bambara (Niger-Congo (Western Mande): Mali) definiteness marker, consisting of a final low tone only (ˋ). If this is then applied to a word where the final tone is high, the tone becomes high-low (ˆ), cf. *káfé* 'coffee' but *káfê* 'the coffee' (Schachter & Shopen 2007: 40).

While tone is a paradigmatic feature in the same way as segments are, variation in loudness, producing **stress** (or **accent**)[35], is a syntagmatic feature in that the contrast is the relationship between adjacent units. With lexical stress (or accent) the contrast is between the syllables of a word. Languages very often distinguish between primary stress (more prominent) and secondary stress (less prominent), as in *ˌphoneˈtician*, where [ˌ] marks the less prominent secondary stress and [ˈ] marks the more prominent primary stress. For a thorough discussion on the differences between tone, stress and pitch-accent, see for example Hyman (2006).

Larger prosodic units, **Intonation Units** (IUs), work on the phrase and the sentence level and usually have the three main functions of "(1) sentence modality and speaker attitude; (2) phrasing and discourse segmentation; and (3) information structure and focus" (Himmelmann & Ladd 2008: 250). Intonation will not be discussed further in this chapter, but rather in the sections discussing the above-mentioned functions.

4.2 Phonological typology

It is an absolute universal that all spoken languages have consonants and vowels. It is also an absolute universal that languages arrange their consonants and vowels into larger units, syllables, which in turn may be arranged into even larger units, words. Furthermore, all languages make use of pitch and loudness one way or another. The following sections give an overview on how languages tend to behave in terms of (segmental) phonemes, syllables and prosody. For an accessible article-length overview of phonological typology, see Maddieson (2010).

35. Strictly speaking, stress is a phonetic parameter while accent is a phonological feature (see, e.g. Himmelmann & Ladd 2008).

4.2.1 Segment inventories

The average number of phonemes, i.e. consonants and vowels, for the languages of the world in Maddieson's (2011b, 2011h) sample is something like 19–25 consonants and something like five to six vowels[36], making the most common phoneme inventory between 24 and 31 contrastive segments. It should be noted here that the vowel inventory in Maddieson (2011h) is restricted to vowel quality only. As we saw above, languages also contrast vowels in terms of length, nasalization and diphthongization. It could, however, be argued that these are not basic features of a vowel inventory. A long vowel could be argued to simply be an instance of two short vowels, and a diphthong to be an instance of vowel-1 plus vowel-2. For instance, Anderson & Otsuka (2006) consider the Tongan (Austronesian (Oceanic): Tonga) long vowels and diphthongs as sequences of two syllables. It could also be argued that nasalized vowels actually consist of a vowel plus a nasal, which is Férère's (1983: 78) analysis of Haitian Creole (Creole (French-lexified): Haiti) as well as Narang & Becker's (1971) analysis of 'Hindi-Urdu'[37]. Thus descriptions of individual languages may turn out to differ considerably. For instance, a language with three vowels, say /a, i, u/, which are contrastively both long and short, where the short vowels can be nasalized, and which has six diphthongs, may be said to have as few as three vowels only (/a, i, u/) or as many as 15 (/a, a:, ã, ai, au, i, i:, ĩ, ia, iu, u, u:, ũ, ua, ui/), or anything in between, depending on how the author analyses length, nasalization and diphthongation. To control for this kind of variation in analyses, Maddieson (2011h) only differentiated vowels according to their quality, i.e. their height, their front-back position, and their roundedness.

Of the 451 languages in the UPSID, the 25 most common consonants and six most common vowels, sorted by frequency, are: /m, k, j, p, w, b, h, g, ŋ, ʔ, n, s, ʧ, ʃ, t, f, l, ɲ, d, ʤ, t̪, kʰ, pʰ, r, v/ and /i, a, u, ɛ, , o, e/.[38] English, counting 24 of them, is an example of

---

36. But note that "far more languages have five or seven vowels than have four or six" (Ladefoged 2005: 37), i.e. have an odd number of vowels. For a discussion on how Maddieson has determined which segments to include in his inventories, and which segments to treat as allophones (thus not including them), see Maddieson (2011b).

37. With 'Hindi-Urdu' the authors mean "the 'common core' language generally spoken by educated Hindus and Muslims" (Narang & Becker 1971: 646) in predominantly the urban areas of roughly the northern half of the South Asian subcontinent. Both Hindi and Urdu are Indo-European Indic languages spoken in India and Pakistan respectively.

38. I have, in fact, on four occasions merged the so-called dental/alveolars (see footnote 32 above) with alveolars, i.e. with /n, t, l, s/. In essence, this means that I have listed the 29 most common consonants of the UPSID, with four of them not individually represented.

Note that these segments are listed according to how many languages have them, which does not necessarily correlate with how frequent the individual sound is in a given language. For instance, in

a language with an average amount of consonants: /p, b, t, d, k, g, f, v, θ, ð, s, z, ʃ, ʒ, ʧ, ʤ, m, n, ŋ, l, ɹ, w, j, h/. The fact that a language has an average number of consonants does not automatically mean that those consonants are the most common consonants among the languages of the world. In fact, we will later see that English contains some cross-linguistically rare consonants.

Table 4.1 gives an overview of both the most common consonants of the UPSID and those found in English. The 19 English consonants that overlap with the most common ones of the UPSID are in a bold outline, while the five English consonants that do not belong to the most common ones are in italics and slightly shaded.

Table 4.1 The 25 most common consonants of the UPSID and the English consonant inventory.

	B		LD		D		A		PA		P		V		G		LV	
Pl	**p**	**b**			t̪		**t**	**d**					**k**	**g**	ʔ			
AspPl	pʰ												kʰ					
Ns		**m**						**n**				ɲ		**ŋ**				
Tr								r										
Fr			**f**	**v**	*θ*	*ð*	**s**	*z*	*ʃ*	*ʒ*					**h**			
Affr									*ʧ*	*ʤ*								
App								*ɹ*				**j**						**w**
LatApp								**l**										

Abbreviations: B = Bilabial; LD = Labiodental; D = Dental; A = Alveolar; PA = Postalveolar; P = Palatal; V = Velar; G = Glottal; LV = Labiovelar; Pl= Plosive; AspPl = Aspirated plosive; Ns = Nasal; Tr = Trill; Fr = Fricative; Affr = Affricate; App = Approximant; LatApp = Lateral approximant

The smallest phoneme inventories in Maddieson's (2011b) sample are found in Rotokas and Pirahã (Mura (Mura): Brazil). Rotokas has only six consonants, /p, t, k, β, ɾ, g/, making it the language with the smallest known consonant inventory, and five vowels /i, e, a, o, u/. Pirahã has eight consonants /p, t, k, ʔ, b, g, s, h/ and three vowels /i, o, a/.[39] Rotokas not only exhibits a rare pattern with respect to the small number of its

the UPSID there are 393 languages that have /i/ and 392 languages with /a/, but the latter tends to be the most frequent vowel in most languages (Ladefoged 2005: 176).

39. Pirahã also has two very uncommon consonants: the first one is the exceptionally rare "voiced, lateralized apical-alveolar/sublaminal-labial double flap with egressive lung air. In the formation of this sound the tongue tip first touches the alveolar ridge and then comes out of the mouth, almost touching the upper chin as the underblade of the tongue touches the lower lip" (Everett 1982: 94). The second is the cross-linguistically rare, but not unheard of, bilabial trill. Both of these sounds, however, tend to occur only in "special types of speech performance" (Ladefoged & Maddieson 1996: 19) and are therefore not included in the general consonant inventory.

consonantal system, but also in the fact that there is no nasal. The fact that a language has a small number of consonants does not automatically mean that its inventory will consist of the cross-linguistically most common ones: two of the consonants in Rotokas, i.e. a third of the inventory, /β/ and /ɾ/, are uncommon, found respectively in 54 (12%) and 7 (1.6%) of the UPSID languages.

While there is general agreement that !Xóõ (Khoisan (Southern Khoisan): Botswana) has the largest phoneme inventory known, the number given differs considerably. In the UPSID !Xóõ is listed as having 141 contrastive segments, 95 consonants and 46 vowels. However, using the criteria of Maddieson (2011h) mentioned above, !Xóõ only has five vowels. On the other hand Maddieson (2011b citing Traill 1985 and 1994) states that !Xóõ has 122 consonants. This would bring the total down to 126 phonemes, which is still a sizeable inventory.

Pirahã is a case in point that a small consonant inventory does not automatically mean a large vowel inventory. In fact, Maddieson states that "absolutely no correlation was found between the number of vowels and the number of consonants" (2011h). He has four languages (0.7%) in his sample with only two vowel quality contrasts: Abkhaz (Northwest Caucasian (Northwest Caucasian): Georgia), Margi (Afro-Asiatic (Biu-Mandara): Nigeria), Mparntwe Arrernte (Australian (Pama-Nyungan): Australia) and Yimas (Lower Sepik-Ramu (Lower Sepik): Papua New Guinea). Mparntwe Arrernte, for instance, has /ə, a/ (Ladefoged & Maddieson 1996: 286). Notice that even with this small inventory, Mparntwe Arrernte makes use of vowel height as a contrastive feature. This is, in fact, another absolute universal (Hyman 2008). As Ladefoged & Maddieson put it, "[a]ll languages have some variations in vowel quality that indicate contrasts in the vowel height dimension. Even if a language has only two phonologically contrastive vowels, the differences will always be in this dimension rather than the front-back dimension" (1996: 286).

Table 4.2 below gives some examples of how different vowel systems may typically look. In each of the systems, the vowels that overlap with the six most common ones of the UPSID listed above have been marked in bold.

What is primarily typical about the systems in Table 4.2 is first of all that height is a contrastive feature in each of them, even in the very small inventories. Secondly, vowel systems tend to be peripherally symmetrical, or, if asymmetrical, as in the case of Quapaw (Siouan (Siouan): USA), they tend to have more front than back vowels (Schwartz et al. 1997). The term 'peripherally symmetrical' is adoped from Schwartz et al. (1997) and means that the vowels located at the outer borders of the IPA chart, such as /i, e, ɛ, a, ɔ, o, u/, tend to occur in the same heights both at the front and the back. It should be noted that /a/ is often part of a vowel inventory, making the typical vowel system roughly triangular. In other words, languages tend "to prefer an odd number of peripheral vowels (hence, one 'low' and an equal number of front

Table 4.2 Some typical vowel systems.

ə **a** 2 vowels: Mparntwe Arrernte (Australian (Pama-Nyungan): Australia) (Ladefoged & Maddieson 1996: 286)	**i** **u** **a** 3 vowels: Standard Modern Standard Arabic (Afro-Asiatic (Semitic): North Africa & Middle East) (Thelwall & Sa'Adeddin 1990)	**i** **e** **o** **a** 4 vowels: Quapaw (Siouan (Siouan): USA) (Rankin 1982)
i **u** **e** **o** **a** 5 vowels: Modern Hebrew (Afro-Asiatic (Semitic): Israel) (Laufer 1990)	**i** **u** **e** **o** ə **a** 6 vowels: Fyem (Niger-Congo (Platoid): Nigeria) (Blench 2006)	**i** **u** **e** **o** **ɛ** **ɔ** **a** 7 vowels: Galician (Indo-European (Romance): Spain) (Regueira 1996)
	i **u** **e** **o** ə **ɛ** **ɔ** **a** 8 vowels: Slovene (Indo-European (Slavic): Slovenia) (Šuštaršič et al. 1995)	

and back ones) whatever the overall number of vowels in a system" (Schwartz et al. 1997: 243). Thus, a system with an even number of vowels, such as Fyem (Niger-Congo (Platoid): Nigeria) with six vowels, will typically have the same five vowels as a 5-vowel system, with the addition of a central vowel, more commonly /ə/ (a pattern found in 24, or 40%, of the 60 languages investigated by Schwartz et al. (1997)[40]. The exception to this pattern, however, is the 4-vowel system, which tends to be asymmetrical. Of the UPSID languages investigated in Schwartz et al. (1997) with a 4-vowel system (25 in total), all were asymmetrical, with the majority (14 or 56%) having more front than back vowels.

40. But cf. Spencer (1996: 120) stating that 6-vowel systems of the type /i, e, a, o, u, ɨ/ are more common than the type /i, e, a, o, u, ə/.

The languages with the most vowels in Maddieson's (2011h) database are German, with 15, and (British) English and Bété (Niger-Congo (Kru): Ivory Coast), each of which has 13. Table 4.3 gives an overview of both the most common vowels of the UPSID and those found in German. The six German vowels that overlap with the most common ones of the UPSID are in a bold outline, while the four German vowels that are cross-linguistically rare are in italics and slightly shaded. For a more accurate placement of each vowel, see the IPA chart in Figure 4.1.

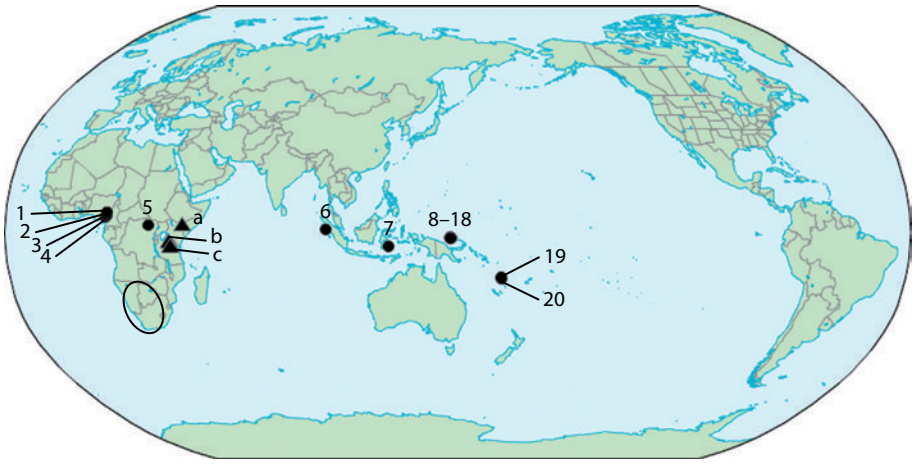
Table 4.3 The seven most common vowels of the UPSID and the German vowel inventory (Kohler 1990).

	Front		**Near-Front**		**Central**		**Near-Back**		**Back**	
	UR	R	UR	R	UR	R	UR	R	UR	R
High	**i**	*y*								**u**
Semi-high			ɪ	*ʏ*				ʊ		
High-mid	**e**	*ø*								**o**
Mid					ə					
Low-mid	**ɛ**	*œ*								ɔ
Semi-low					ɐ					
Low			**a**							

Abbreviations: UR = unrounded; R = rounded

We have seen which the most common consonants and vowels are. While it seems intuitively straightforward that a large inventory will include uncommon phonemes, we have already seen in Rotokas that a small inventory does not necessarily include only common phonemes. It is also important to keep in mind that simply because we are used to certain phonemes, because they occur in well-known languages, it does not follow that these are common cross-linguistically. A case in point are the dental non-sibilant fricatives of English /θ, ð/, which occur in only 40 of 567 languages, i.e. 7.1%, in Maddieson's (2011d) database. Other rare consonants include clicks, found in 9 (1.6%) of the languages in Maddieson's (2011d) sample and only in Africa. All Khoisan languages have them, but they are also found in other southern and eastern African languages, albeit with a much lower frequency and with a much smaller proportion of words containing clicks, such as Zulu (Niger-Congo (Bantoid): South Africa) and Dahalo (Afro-Asiatic (Southern Cushitic): Kenya) (Ladefoged & Maddieson 1996: 246). The fact that clicks occur as far north as Hadza (Isolate) and Sandawe (Khoisan (Sandawe)) in Tanzania, as well as Dahalo in Kenya, is one of the reasons for the assumption that languages with clicks were once more widespread. The circle on Map 4.1 indicates the area where most click languages are found, while the three triangles mark Hadza, Sandawe and Dahalo.

A consonant that was until relatively recently thought non-existent (cf. Dryer 1997: 124) is the bilabial trill /ʙ/, i.e. a trill (see above) formed with both lips, which is found as an ordinary phoneme in a handful of languages in Africa (predominantly West Africa) as well as a number of Austronesian languages (Maddieson 1989), as shown on Map 4.1. Languages that have bilabial trills as allophones (for example Pirahã, as mentioned above), or only in a limited set of the lexicon, have not been indicated on the map.



Map 4.1 Black dots indicate languages with a bilabial trill phoneme. The circle indicates the area where clicks are typically found, and the triangles show three click languages north of the typical click language area. Legend (unless otherwise mentioned, from Maddieson 1989): 1. Icen (Niger-Congo (Jukunoid): Nigeria; Roger Blench, p.c.), 2. Kom, 3. Babanki, 4. Ngwe, all Niger-Congo Bantoid languages in Cameroon, 5. Mangbetu (Nilo-Saharan (Mangbetu): DR Congo; (Demolin 1991)), 6. Nias (Austronesian (Sumatra): Indonesia), 7. Muna (Austronesian (Sulawesi): Indonesia), 8. Ponam, 9. Ahus (Hus), 10. Leipon, 11. Kurti, 12. Koro, 13. Lele, 14. Papitalai, 15. Kele, 16. Nali, 17. Ere, 18. Titan, all Austronesian Oceanic languages in Papua New Guinea, 19. Uripiv, 20. Naʔahai, both Austronesian Oceanic languages in Vanuatu; a. Dahalo, b. Hadza, c. Sandawe.

An extremely rare consonant among the languages of the world is the interdental approximant, found in, for example, the Californian dialect of American English. Here the approximant for the word *this*, pronounced /ðis/ in most English varieties, is formed by letting the tongue protrude between the teeth in such a way that the tongue blade moves towards the upper teeth without closing off the air passage /ð̞/. In some languages, such as Bauchi Guda (Niger-Congo (Kainji): Nigeria), the interdental is

phonemic. In fact, Bauchi Guda has at least three contrastive interdental approximants: /ð̪/, /ð̪ʷ/ and /ð̪ʲ/ (Harley & Blench fc: 10).

Because many of us are familiar with European languages such as French and German, we might not be aware of the fact that the high and mid front rounded vowels /y, ʏ, ø, œ/ common in those languages are in fact cross-linguistically rare.[41] Only 6.6% in Maddieson's (2011c) sample have them. Most of those languages can be found in northern Eurasia, though there is one in Africa, Ejagham (Niger-Congo (Bantoid): Nigeria), three in the Pacific region, Malakmalak (Australian (Northern Daly): Australia), Iaai (Austronesian (Oceanic): New Caledonia) and Natügu (Austronesian (Oceanic): Solomon Islands), and four on the American continent, Hopi (Uto-Aztecan (Hopi): USA), Quiotepe Chinantec (Oto-Manguean (Chinantecan): Mexico), Wari' and Aikaná (Isolate: Brazil)[42]. In general, those languages that have front rounded vowels tend to have a larger than average vowel inventory (Maddieson 2011c).

Another example of a vowel we are rather used to, but which is cross-linguistically very rare, is the rhotacized, lower-mid central unrounded vowel /ɝ/, which is very common in American English (for example in the word *bird*) and can also be found, though less frequently, in Mandarin (Ladefoged 2005: 29). A similar vowel, /ɚ/, can be found in Changzhou Wu (Sino-Tibetan (Chinese): China) (Chao 1970: 48) as well as in Yurok (Algic (Yurok): USA) (Blevins 2003).

There are so many known phonemes that are extremely rare, occurring in very few or even only one language, such as the Swedish /ɧ/ (usually described as a simultaneous ʃ and x)[43], that it would be impossible to list them all. In fact, about 46% of all the segments in the UPSID occur only in one single language. According to Ladefoged (2005: xiii) there are about 200 different vowels and more than 600 different consonants in the world's languages.

---

41. Swedish actually has two high front rounded vowels, /y/ and what I will denote as /ʉ̟/ (following Engstrand 1990: 43), by which I mean that the sound is a front and not a central vowel. The two vowels differ in that /y/ has what Ladefoged & Maddieson (1996: 296) term "(horizontal) lip rounding and protrusion" ('outrounded' in Engstrand 1990: 43) while the latter has "(vertical) lip compression" ('inrounded' in Engstrand 1990: 43). Swedish thus has a minimal set, *ni* /ni:/ 'you (pl)', *ny* /ny:/ 'new' and *nu* /nʉ̟:/ 'now', where the only contrastive feature is the (type of) lip rounding.

42. This classification follows Anonby (2009) and Fabre (2009), among others, and differs from the WALS classification which lists Aikaná as an Arawakan language.

43. The symbol /ɧ/ actually covers several, slightly different sounds in Swedish (Elert 1989: 76f, for example, differentiates between three). However, the variation is regional and allophonic, not phonemic.

BACK UNROUNDED VOWELS

Back unrounded vowels, /ɯ, ɣ, ʌ, ɑ/, are also rather uncommon; Maddieson states that "front rounded vowels are more frequently found than back unrounded" (1984:125). However, a query in the UPSID for front rounded monophthongs yields 32 (7.1% of 451) languages, while a query for back unrounded monophthongs yields 85 (18.9%) languages with at least one back unrounded vowel. This discrepancy could possibly be due to the fact that the UPSID only contained 317 languages in 1984, while it now contains 451. Of these, 47 languages have /ɯ/, a sound often thought of as 'exotic' because it rarely occurs in Europe – though Scots Gaelic (Indo-European (Celtic): UK), which is not in the UPSID, has it (Ladefoged et al. 1998), as does Turkish. (Actually, Scots Gaelic has both /ɯ/ and /ɣ/, which places it in the group of languages that has the more common type of 2-way contrast.) Again, because we are so familiar with English, we might not be aware of the fact that the sound /ʌ/, as in the word *luck*, is in fact quite rare. It is found in only 10 (2.4%) of the languages of the UPSID. Figure 4.2 gives a summary of the back unrounded vowels in the UPSID plus English.

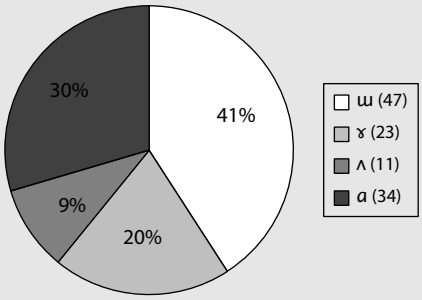


Figure 4.2 Proportions of back unrounded vowels in UPSID with English added.

21 languages, including English, have either two (15) or three (6) back unrounded vowels. English, with /ʌ, ɑ/, and Nishi (Sino-Tibetan (Tani): India), with /ɯ, ʌ, ɑ/, are the only languages that contrast the two low ones. The most frequent 2-way contrast is /ɯ, ɣ/, found in eight UPSID languages. (But note that Irish (Indo-European (Celtic): Ireland) is listed in the UPSID as having only one back unrounded vowel, /ɑ/. Ní Chasaide (1999), however, lists /ɯ/ and /ɣ/ as part of the Irish vowel inventory.) Vietnamese (Austro-Asiatic (Viet-Muong): Vietnam) is an example of a language with a 3-way contrast, /ɯ, ɣ, ʌ/.

4.2.2 Patterns of syllable structures

Almost all languages allow syllables of CV, where the onset C stands for any consonant and the nucleus V for any vowel (of any length, monophthong or diphthong). A very rare exception to this near-absolute universal can be found in four dialects of Kunjen (Australian (Pama-Nyungan): Australia), namely Oykangand, Olgol, Okunjan and Kawarrangg (Sommer 1970). Most languages allow syllables consisting only of a

vowel (V), or consisting of two consonants and a vowel (CCV or CVC). Languages that allow anything up to the latter two types are labelled as having moderately complex syllables in Maddieson (2011e), and make up more than half (56.4%) of his database. Most often the second consonant in the CCV syllables can only be a liquid or a glide. For instance, the only consonant clusters allowed in Chamorro (Austronesian (Chamorro): Guam) are Cl, Cr and Cw (Topping 1973: 36f);[44] consequently a word like *dop.bla* 'bend' (the dot . indicates a syllable break) contains the two most complex types of syllables allowed in the language.

Some languages, however, only allow simple (C)V syllables, where (C) means that the onset is optional. An example of this would be Yoruba (Niger-Congo (Defoid): Nigeria), where *a.wo* 'cult' illustrates the only types of syllables allowed (Bamgbose 1966: 6, 9). There are also languages that do not permit syllables consisting only of a nucleus, such as Mantjiltjara (Australian (Pama-Nyungan): Australia), which has only two types of syllables, CV and CVC (Marsh 1969: 145), as in *ju:* 'yes', *ma.ji* 'food', *wil.tu* 'hard' and *ŋa.kum.pa* 'deaf'.[45] And while it is rare, there is at least one known language, Hua (Trans-New Guinea (Eastern Highlands): Papua New Guinea), where the only type of syllable allowed is CV (Haiman 1980a).[46] Languages that only allow the simple syllables (C)V make up 12.5% of Maddieson's database.

The middle group of languages is that allowing complex syllables, i.e. syllables with more than three consonants plus a vowel. This is found in 31.1% languages of the sample. English is an example of a language with complex syllables, allowing a consonant cluster of up to three in the onset and up to four in the coda, CCCVCCCC, although this kind of syllable is rather infrequent in the language (the example typically given is *strengths* if pronounced [strɛŋkθs]). However, words with either three consonants in the onset or four in the coda are not infrequent, such as *scrubs* [skrʌbz] and *glimpsed* [glɪmpst]. These words are also examples of the sonority principle mentioned above. Georgian (Kartvelian (Kartvelian): Georgia) is an example of a language that allows very complex syllables indeed, with up to six consonants in a cluster: *mc'vrtneli*[47] 'trainer' (Butskhrikidze 2002: 106). Languages that allow complex syllables also tend to allow simple syllables.

Maddieson (2011e) found a certain correlation between syllable structure and consonant inventories: languages with simple syllable structures tend to have smaller

---

44. The first two types of clusters, Cl and Cr, only occur in words borrowed from Spanish or English. For more on borrowing and language contact, see 13.2.

45. It's slightly more complicated than that, though, as the language allows CCV and C suffixes.

46. This in fact only represents the underlying structure of the syllables; other structures may occur, under specific conditions. See further Haiman (1980a).

47. *c'* is one single consonant and stands for a glottalized, or ejective, alveolar affricate [ʦ'].

consonant inventories, and, conversely, languages with complex syllable structures tend to have larger consonant inventories. However, he goes on to say that "these overlaps might be due to fortuitous distribution of genealogically-shared or areally-spread features" (ibid.).

4.2.3 Patterns of prosody

4.2.3.1 *Length*

All spoken languages have short vowels and consonants. It is also not uncommon for languages to have long vowels, though the UPSID only lists 51 languages, or 11.3% of the database, containing them. It should be noted here that that figure covers languages which have any kind of long vowels, irrespective of whether they contrast with short vowels of the same quality. For instance, English has a number of long vowels, but none of them contrast with a short vowel of the same quality. Thus /i:/ is basically the long counterpart of /ɪ/ in the pair *heal*/*hill*, /u:/ is the long counterpart of /ʊ/ in the pair *fool*/*full*, and so on. An example of a language where length is truly a contrastive feature, i.e. where the same vowel quality can occur either as a short or a long vowel, is Croatian, as shown in (17) (I am, for the moment, ignoring the pitch accent distinctions of the language).

Croatian (Indo-European (Slavic): Croatia)

(17)

	short	**gloss**	**long**	**gloss**
/i/	vile	'hayfork'	vi:le	'fairies'
/e/	tek	'only'	te:k	'appetite'
/a/	pas	'dog'	pa:s	'belt'
/o/	kod	'by, at'	ko:d	'code'
/u/	duga	'stave'	du:ga	'rainbow'

(Landau et al. 1995: 84)

Languages can even have three length distinctions for vowels. The UPSID lists 9 (2%) languages with so-called reduced (or extra-short) vowels. Khanty (Uralic (Ugric): Russia), for instance, has contrastive full, long and reduced vowels, as in *köt* /køt/ 'hand' contrasting with reduced vowel in *kŏt* /kŏt/ 'distance' (Gulya 1966: 24). Lanyjang Dinka (Nilo-Saharan (Nilotic): Sudan), on the other hand, is an example of a language with short, long and overlong vowels, as in /láŋ/ 'kind of berry' (PL) contrasting with /làaŋ/ 'kind of berry' (SG) and /lăaaŋ/ 'slave' (SG) (ignoring, for the moment, the differences in tone; Remijsen & Adong Manyang 2009).

Languages may also have long consonants, and the UPSID lists 12 (2.7%) such languages. An example of a language with contrastive consonant length is Central Berawan (Austronesian (Northwest Malayo-Polynesian): Malaysia) (specifically the Long Teru dialect), with pairs such as /liŋŋen/ 'self' versus /liŋen/ 'out of sight' (Blust 1995: 128). Some languages have three consonant lengths, such as Inari Saami (Uralic

(Finnic): Finland) with short, half-long and long consonants, though the latter two are usually not distinguished in orthography anymore. Compare, for instance, various grammatical cases of the word *pínnoo* 'stack, pile': NOM *pínnoo* /pín'o:/ (half-long *n*), GEN/ACC pínoo /píno:/ (short *n*), ESS[48] *pínnòon* /pín:ò:n/ (long *n*) (Bye 2007: 62).[49]

4.2.3.2 *Tone*

Tone is a very common feature in the languages of the world, especially in Africa and Southeast Asia. However, in Maddieson's (2011f) sample the non-tonal languages dominate.

> This probably underrepresents the proportion of the world's languages which are tonal since the sample is not proportional to the density of languages in different areas. For example, (…) less than 5% of the Niger-Congo languages are included. (…) <But> over 10% of the Western European languages listed are included, only two of which are tonal or marginally so and the rest non-tonal. If, correspondingly, 10% of the Niger-Congo family had been included, 80 additional tone languages would have been included.
>
> (Maddieson 2011f)

The tonal languages in Maddieson's sample were coded for whether they had simple tone, contrasting only two tones (usually high and low), or whether they had a more complex tone inventory. Simple tone languages, making up a quarter of the total database, are more frequent than complex tone languages (with 16.7%). There are almost no tonal languages in Europe, though Norwegian (Indo-European (Germanic): Norway), Swedish, Latvian (Indo-European (Baltic): Latvia; (Brenzinger 1973)) and Goizueta Basque (Isolate: Spain; Hualde et al. 2008) are exceptions, all with simple tone systems.[50] Languages in South Asia, the islands of Southeast Asia (such as Indonesia and the Philippines) and Australia tend not to have tones. While there are tonal languages on the American continent, the majority seem to be non-tonal. Tone languages are very common in Africa and continental Southeast Asia, with complex systems dominating in the latter area. Tilquiapan Zapotec (Oto-Manguean (Zapotecan): Mexico) is an example of a language with a simple tone system, having only two contrastive tones,

---

48. ESSIVE case can be roughly translated into 'as something', in this example the essive would translate into something like 'as a pile' or 'as a stack'.

49. Long consonants, sometimes also called over-long (as opposed to the long, which are here labelled half-long), were previously distinguished orthographically from half-long consonants by an apostrophe, which in this case would mean a spelling like pínoo – pínnoo – pín'noon (Trosterud & Uibo 2005: 142).

50. Goizueta Basque not only has two contrastive tones, but also two contrastive types of stress, effectively giving a 4-way contrast.

high and low, as in *biaxtily* /biaʃtíli/ 'soap' versus *biaxtily* /biaʃtìli/ 'white prickly pear' (Merrill 2008:111). An example of a language with a complex tone system is Eastern Kayah Li, contrasting three level plus two contour tones, as in Table 4.4.

Table 4.4 The tone system of Eastern Kayah Li (Sino-Tibetan (Karen): Myanmar). (Adapted from Solnit 1997:20)

	Tone symbol	Tone number[51]	Tone description	Gloss
kō	꜖	33	mid level	blow away
ko	ꜘ	11	extra low level	(general classifier)
kò	ꜗ	21	low falling	wear on head
kó	ꜙ	55	extra high level	do temporarily
chiko	ꜛ	52	high falling	shrimp

There are no known languages with more than five level tones (Blevins 2001b). Gimira is an example of a language that has five level tones plus one contour tone, giving a minimal set of up to six contrasts, as in Table 4.5:

Table 4.5 The tone system of Gimira (Afro-Asiatic (North Omotic): Ethiopia). (Rapold 2006:120)

	Tone symbol	Tone number	Tone description	Gloss
kə̀r	ꜘ	11	extra low level	loincloth
kàr	ꜗ	22	low level	wasp
kə̀̄r	꜖	23	low mid	game with stones
kār	꜖	33	mid level	to circle
kári	ꜙ	44	high level	inset; banana leaf
kə́r	ꜙ	55	extra high level	clear

Ticuna (Isolate: Brazil) is another example of a language with a very complex tone system, with five level tones, as well as four contour tones (Anderson 1962; cf. also Soares 1992).

Interestingly, the languages in APiCS exhibit a slightly different pattern (Michaelis et al. 2013: feature 120). The majority are non-tonal, which is parallel to the languages in Maddieson's WALS database, but the ratio of non-tonal languages is higher among

51. Tone numbers are used to indicate the movement of the tone during the utterance of the TBU. Level tones start and end at the same place, for instance at the mid level (3), and is therefore coded as 33 (which means that the tone moves from level 3 to level 3). Contour tones move and the number indicates the direction by coding the start and end points. Thus a 21 tone means that the tone started at level 2 and fell to level 1.

the languages in APiCS (52 of 74[52] languages or 70.3% in APiCS to the 307/526 or 58.4% in WALS after the pidgins and creoles have been subtracted from Maddieson's sample). The proportion of APiCS languages with a simple tone system (21 languages or 28.4%) is slightly higher than the proportion of non-creoles with simple tone system. However, only one (1.4%) APiCS language, Sango (Creole (Ngbandi-lexified): Central African Republic)[53] has a complex tone system, which is of a much lower proportion than the 16.7% of the WALS languages. Harking back to the fact that the tone languages were probably underrepresented in WALS and that the APiCS languages had a higher proportion of non-tonal languages, this seems to imply that (i) pidgin and creole languages are less likely to be tonal, and (ii) if they are, it is less likely that pidgin and creole languages will have a complex tone system than a non-creole language.

Maddieson found that "tonal complexity and complexity of segment inventory tend to go hand in hand across the set of languages surveyed" (2011f). In his sample, the languages with complex tone systems also tended to have larger inventories of both consonants and vowels. The correlation was inverse for syllable structures, in that languages with complex tone systems tended to have less complex syllable structures, while languages with complex syllable structures tended to not be tonal. Put differently, the more complex the syllable, the less likely that it will be a tonal language.

4.2.3.3 *Stress*

A little over half (56.2% to be precise) of the 502 languages surveyed by Goedemans & van der Hulst (2011a) have fixed stress. Very simplified, languages with fixed stress can be divided into two major groups: those where primary stress is placed relative to the left end of the word and those where primary stress is placed relative to the right end of the word. I will here arbitrarily assign the first group the label LS (for Left Stress) and the second RS (for Right Stress) for the sake of being able to have an easy shortcut to distinguish the two in the present section. However, it should be kept in mind that these labels are not used in studies dealing with stress and rhythm. Note also that I will not discuss patterns of secondary stress.

The majority of the languages with fixed stress in Goedemans & van der Hulst's sample are RS languages (61.3%) though LS languages are also common (38.7%) and both groups are spread over the world although there seems to be a slight tendency

---

52. Since Singlish is marked as having a complex tone system of only marginal frequency (Ansaldo & Lim 2013), I have included it in the group of languages marked as non-tonal. Information is not available for Batavia Creole (Creole (Portuguese-lexified): Indonesia) and Early Sranan, which is why the figure of total languages is 74 for tone, despite the fact that the total number of languages in APiCS is 76.

53. Note that this classification follows that of APiCS and not that of WALS, where Sango is classified as a Niger-Congo Ubangi language.

for LS in European and Australian languages and a tendency for RS in Austronesian languages. For each group, stress can be located at one of three positions relative to the end of the word: either the syllable at the end, or one syllable in from the end, or two syllables in from the end. For LS that means that stress is either located on the first syllable ('**initial** stress'), the second syllable ('**second** stress') or the third syllable ('**third** stress') of the word. Initial stress is by far the most common strategy, with 92 (84.4%) languages. An example of a language with initial stress is Kayardild (Australian (Tangkic): Australia), as in ˈ*malaa* 'sea' (Evans 1995: 81).

A minority of the LS languages have the stress on the second syllable, such as Dakota (Siouan (Siouan): USA) *waˈkte* 'I kill' (Shaw 1985: 175). Only one single LS language in the database, Winnebago (Siouan (Siouan): USA), regularly places stress on the third syllable from the left, as in *hopiˈrak* 'belt' *haraˈčabra* 'the taste' (Miner 1979: 28).

The three types of stress mapped for RS languages are the last syllable ('**ultimate** stress'), the second to the last syllable ('**penultimate** stress') and the third to the last syllable ('**antepenultimate** stress') of the word. The majority (63.6%) of the RS languages place their stress on the penultimate, such as Rapanui (Austronesian (Oceanic): Easter Island), for example *maˈneŋe* 'small' (Du Feu 1996: 193). The second largest group (29.5%) of the RS languages place their stress on the ultimate. An example of a language with consistent ultimate stress is Bashkir (Altaic (Turkic): Russia), as in /kitˈap/ 'book' and /kitapˈlar/ books (Poppe 1964: 18). Antepenultimate languages comprise the smallest group (6.9%). Georgian, for instance, has an antepenultimate stress pattern, as in *dedali* /ˈdɛdalɪ/ 'hen' (Butskhrikidze 2002: 97).

220 languages (or 43.8%) in the sample do not have fixed stress, but allow stress to occur on different syllables. English is an example of such a language, as in the pair ˈ***per**mit* (noun) and *per*ˈ***mit*** (verb) or in the set ˈ***de**mocrat*, *de*ˈ***mo**cracy*, *demo*ˈ***cra**tic*. In many languages, English among them, even if stress isn't fixed, the location of the stressed syllable is still consistently either towards the left or the right of the word boundary. It is far more common in Goedemans & van der Hulst's (2011b) sample that the stress occurs somewhere on the right (41.8%), such as in English, but in a fair amount of languages (17.7%) stress is placed somewhere to the left. An example of such a language is Malayalam (Dravidian (Southern Dravidian): India), as in /ˈpukavaɳʈi/ 'train' (primary stress is on the first syllable) and /paʈˈʈaːɭaḵaːram/ 'soldier' (primary stress is on the second syllable) (Asher & Kumari 1997: 436). There are no left-oriented systems in Africa and very few on the Eurasian landmass in Goedemans & van der Hulst's sample. About a quarter of the languages without fixed stress (24.6%) have so-called unbounded systems, meaning that stress can be anywhere. This kind of system is spread all over the world, although it is rare in the Asian part of the Eurasian landmass. An example of a language with an unbounded system is Seri (Hokan (Seri):

Mexico) (Marlett 2008). A few languages (only 8 or 3.6% in the sample) combine this type of unbounded system with a right-oriented system. Danish (Indo-European (Germanic): Denmark) is an example of such a language (Grønnum 1998). Finally, there are languages (10.9%) where the stress location is not predictable, for instance Rama (Chibchan (Rama): Nicaragua) (Grinevald n.y.).

In most of the languages with predictable non-fixed stress, the determining factor of where the stress occurs has to do with **syllable weight**. A syllable can be made heavy either because it has a long vowel, or because it has a coda, or both. In some languages the position of stress is lexically contrastive. For instance, the only difference between the two Greek words /ˈjɛros/ 'old man' and /jɛˈros/ 'strong' (Hionides 2002) or the two Italian (Indo-European (Romance): Italy) words /ˈcantɔ/ 'I sing' and /canˈtɔ / 'I sang' (source: personal knowledge) is whether the stress is on the first or second syllable.

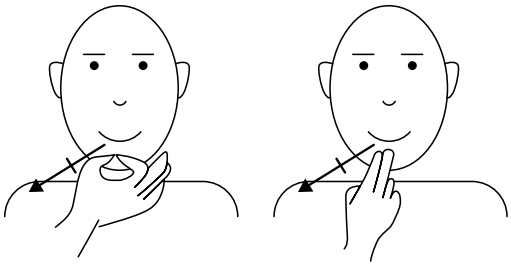
4.3 Sign language phonology

We have seen that spoken languages have a limited set of linguistically contrastive segments, that is, units that are in themselves meaningless, but that can be combined into larger units, syllables, which in turn either in themselves serve as words or can combine into words. This is also the case with signed languages. And while the terminology is borrowed from the study of spoken languages, the fact that sign languages have phonemes and syllables should hardly be a difficult concept. William Stokoe (1960), in his greatly influential analysis of ASL, showed that the phonemes of signed languages are, just as in spoken languages, a limited set of contrastive segments that are built up from a set of distinctive features. These contrastive segments, phonemes, are then combined into larger units, signs. By changing one of the segments, the meaning of the sign is changed – forming minimal pairs or sets.

Signs are formed either with one hand or both, but there is usually one dominant hand, with the other hand serving a subordinate, but significant role, mainly in the domain of prosody and grammar. There is no known sign language where the choice of hand is significant (Emmorey 2002). In other words, signers use the hand they are most comfortable with as the dominant hand.

Stokoe (1960) distinguished between three parameters for a segment: **handshape**, **location** and **movement**.[54] Handshape denotes the shape of the hand during the sign, for instance if the hand is formed as a fist, if the fingers are spread or together, if

54. Stokoe coined the terms **chereme** and **cherology** (from Greek *cheir* 'hand') for the minimal meaning distinguishing units of sign languages, as a parallel to *phoneme* and *phonology* in oral languages. However, the terms have by now largely been abandoned in favour of 'phoneme' and 'phonology'.



(a) **Ask** (b) **Answer**

Figure 4.2 The DGS sign for ASK and ANSWER (source: personal knowledge). Illustration: Maria Johanson. Used with permission.

they are bent, and so on. The location refers to where the hand is on the body of the signer, or around the space of the signer. For instance, the hand can be around the head, the mouth or the chest; it can be directly in front of the signer or to the side, and so on. Finally, movement denotes how the hand moves during the sign, that is, whether it moves away from or towards the signer, up or down, left or right, and so on. Subsequent research has added **hand orientation** and **non-manual** features to the inventory. Hand orientation denotes the way the palm faces during the sign. This feature can, together with the feature handshape, also be considered a subfeature of the **hand configuration** (cf. for example Sandler & Lillo-Martin 2006). Non-manuals include, for example, facial expressions and body posture. Thus sign languages also have minimal pairs, such as in DGS, where the difference between 'ask' and 'answer' lies only in the handshape, as shown in Figure 4.2.

In Figure 4.2a and b the location and movement of the sign are the same, but the handshapes differ, with the index and thumb meeting to form a ring and the three remaining fingers outstretched in the sign for ASK, but with the index and middle fingers outstretched and the rest bunched up in a fist for the sign for ANSWER.

While almost everybody agrees that sign languages also arrange their contrastive segments sequentially to form syllables, the analyses of the exact nature of a syllable differ. Liddell (1984) proposes that syllables consist of two kinds of segments, **hold** (the period of time when none of the components of the sign change) and **movement** (the period of time when any of the components of the sign may change), analogous to consonants and vowels of spoken languages. Another model (see, for example, Corina & Sandler 1993) proposes that it is **location**, rather than hold, which combines with movement to form syllabic sequences. Yet other models include, for example, Perlmutter (1992), where movement combines with **position**, and van der Hulst (1993), who effectively considers movement as only a transition between locations. In many sign languages the majority of the signs are monosyllabic.

The dynamics of the feature movement allows it to be likened to a kind of visual sonority (Sandler 2003) and there are several proposals for a sign language sonority scale analogous to the spoken language sonority scale, among others Sandler (1993) and Brentari (1993).

Sign languages also make use of suprasegments, that is, contrastive units that may carry over across segments, such as tone in spoken languages. These are predominantly non-manuals, but hand configurations can also be considered to belong to this category, since handshape and orientation can be consistent throughout an L-M-L (location-movement-location) sign. Non-manuals may carry such prosodic information as intonation and intonation units, for instance expressing content (*wh*-) questions or shared information (Pfau & Quer 2010). For example, ISL forms polar questions with the non-manual signs of raised brows, widened eyes and the head tilted forwards during the sign (Nespor & Sandler 1999: 171).

Phonological inventories of sign languages differ. ASL, for instance, has 25 handshapes (Tennant & Brown 1998: 28), while the NGT has 31 (van der Kooij 2002: 154ff), AdaSL has 29 (Nyst 2007: 70) and Al-Sayyid Bedouin Sign Language (ABSL) in Israel has 15 (Israel & Sandler 2011). As yet there are no cross-linguistic macro-surveys of the phonological systems of sign languages, although research on individual sign languages is vibrant and growing. Such questions as how many handshapes may be common to all or most of the known and described sign languages are therefore a matter for future research.

4.4 Summary

Both spoken and signed languages systematically make use of a set inventory of linguistically contrastive segments to build larger units of syllables and words.

In spoken languages these segments can be divided into consonants and vowels. Vowels are formed by letting air flow freely through the oral cavity, modified by the height and position of the tongue. Nasal vowels allow air to also flow through the nasal cavity. Consonants are formed by creating various kinds of obstacles for the air flow, partially or completely. Some consonants are formed by sucking air in instead of letting it flow out.

Suprasegmental features such as length (the duration of the production of the segment), tone (modifications in pitch) and stress (modifications in volume of the unit) may also be contrastive.

Languages differ radically in terms of their segment inventories, ranging from 11 to 126 phonemes. The average phoneme inventory has 19–25 consonants and 5–6 vowels. Almost all languages allow their phonemes to combine to (C)V syllables, with optional onsets. Many languages are tonal, most commonly distinguishing between two tones (usually high and low). It is slightly

more common to have fixed stress, usually on the penultimate. Languages that do not have fixed stress tend to place stress on one of the syllables towards the right end of the word.

In signed languages the contrastive segments are either hold and movement or location and movement. These are arranged sequentially to form syllables. Non-manuals also tend to be contrastive, functioning on a suprasegmental and prosodic level. Choice of hand is not contrastive. Further research may reveal cross-linguistic patterns on inventories and structures of the phonology of sign languages.

4.5 Keywords

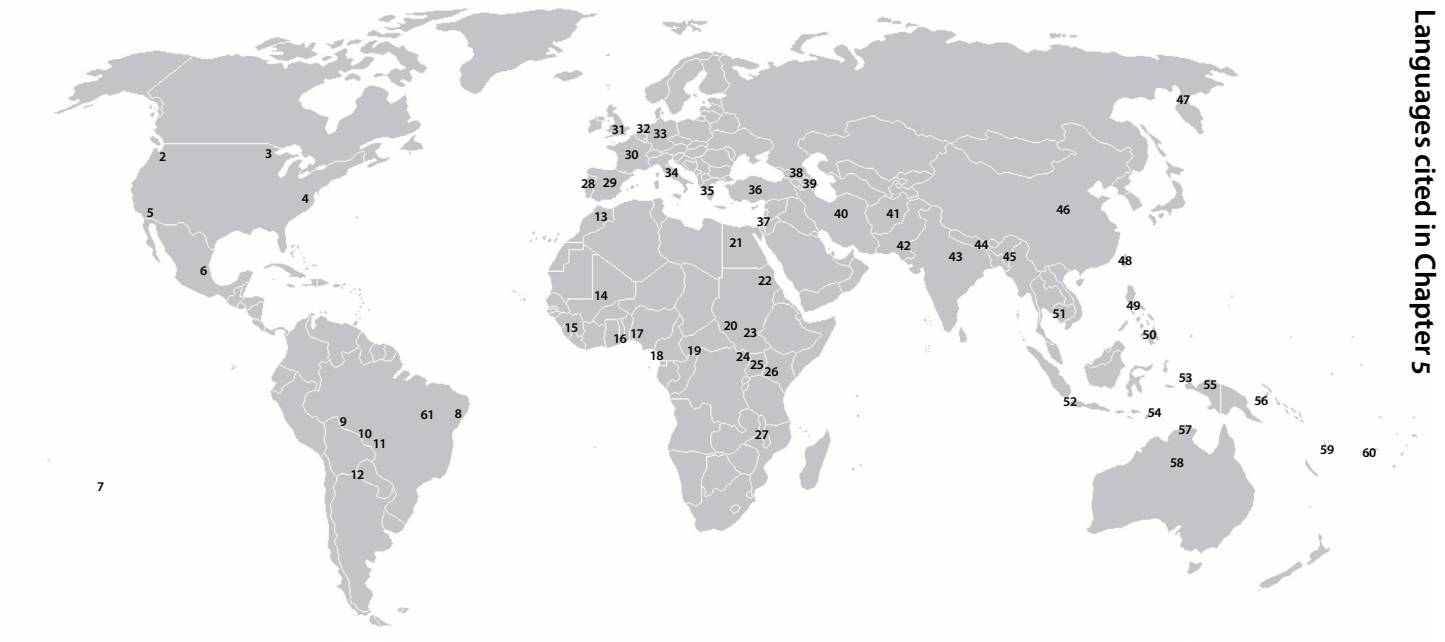
consonant
intonation unit
length
phoneme
phonology
prosody
segment
stress
suprasegment
syllable
tone
vowel

4.6 Exercises

1. What are the contrastive segments of spoken and signed languages respectively?
2. What is the **sonority principle**?
3. Why are **length**, **tone** and **stress** called suprasegmental features?
4. How do patterns of tone, segments and syllables correlate?
5. Is the following statement true or false? Motivate your answer.

   Pidgin and creole languages seem less likely than non-creole languages to have tone.

Languages cited in Chapter 5



1 Hawai'i Creole English
2 Columbia-Wenatchi
3 Ojibwe, Minnesota
4 American Sign Language
5 O'odham
6 Otomí, Sierra
7 Rapanui
8 Kipea
9 Wari'
10 Mamaindê
11 Bororo
12 Wichí
13 Berber, Middle Atlas
14 Koyra Chiini
15 Kisi
16 Logba
17 Yoruba
18 Pichi
19 Sango
20 Dinka, Agar
21 Arabic, Egyptian
22 Beja
23 Nuer
24 Lugbara
25 Lango
26 Maasai
Nandi
27 Chichewa
28 Portuguese
29 Spanish
30 French
31 English
32 Dutch Sign Language
33 German
German Sign Language
34 Italian
Latin
35 Greek
36 Turkish
Turkish Sign Language
37 Hebrew
Israeli Sign Language
38 Ingush
39 Udi
40 Persian
41 Pashto
42 Indo-Pakistani Sign Language, Karachi
43 Indo-Pakistani Sign Language, Delhi
44 Chintang
45 Meithei
46 Mandarin
47 Alutor
48 Thao
49 Chabacano, Caviteño
Chabacano, Ermiteño
Chabacano, Ternateño
Ilocano
Tagalog
50 Maranao
51 Kasong
52 Indonesian
53 Maybrat
54 Leti
55 Iau
56 Nakanai
57 Tiwi
58 Warlpiri
59 Erromangan
60 Fijian
61 Portuguese, Brazilian

Chapter 5

Morphology

We have seen that languages make use of contrastive segments, the smallest units that differentiate between meanings. These small elements do not in themselves convey any meaning, they merely serve contrastively. Both spoken and signed languages make use of another type of minimal element, namely those that do convey some kind of meaning: morphemes. This chapter gives an overview of what kinds of morphemes languages may have and how they may combine. I will first give a brief sketch of the very basics of morphological concepts (5.1). In Section 5.2 I give an overview of how languages tend to employ morphemes. I first discuss patterns on how morphemes combine and stand in relation to each other (5.2.1, 5.2.2 and 5.2.3) and show that the traditional way of classifying languages obscures much of the morphological complexity that languages of the world may have. I then bring up the issue of how much information words may carry (5.2.4). Finally, in Section 5.3 I give an overview of sign language morphology and how it both overlaps with and differs from spoken language morphology.

5.1 The building blocks of words

This section is meant as a very quick overview of the basic concepts of morphology. It is by no means exhaustive. For very accessible introductions to morphology, see, for example, Haspelmath (2002) and Bauer (2003). For a more in-depth overview of the various concepts, including a set of language descriptions, see, for example, Spencer & Zwicky (1998) and Booij et al. (2000).

Morphology is the study of shapes. We have seen that languages make use of small contrastive units that in themselves don't convey any meaning, phonemes. The smallest unit of a language that does convey some kind of information is a **morpheme**. And just as a phoneme is a linguistic abstraction, which can be realized in different ways (allophonically), so a morpheme is an abstraction of the various types of **morphs** that a language has. For instance, in English the concept of plural can be realized in several different ways: as *-s* (cat/cat**s**), *-z* (dog/dog**s** [dɒg**z**]), *-es* (dish/dish**es**), *-en* (ox/ox**en**), *-Ø* (sheep/sheep), or a vowel change in the stem (foot/f**ee**t). All of these individual morphs convey the same piece of information, namely that of 'plural', and as such are **allomorphs**.

Morphemes can be either **free** or **bound**. A free morpheme stands alone as its own word, while a bound morpheme needs some kind of host to attach to. In the sentence

He will go home tomorrow there are only free morphemes, while the sentence *The oxen pulled the chart* contains two bound morphemes, *-en* and *-ed* in *ox**en*** and *pull**ed***. These two bound morphemes, conveying the information of 'plural' and 'past tense' respectively, cannot stand alone as their own words, but have to attach to a host.

The **root** of a word is the smallest unit – and as such is a morpheme – with any semantic content. For instance, the semantic content of HAIR is conveyed through its root *hair*. This root can then be modified, for example by adding the plural marker *-s* to form the word *hairs*. The root of *unhappiness* is *happy*, which has been modified with *un-* and *-ness*. A **stem** is the base for an inflected word form. It can consist minimally of a root, but may also be a modification of the root in some way. For instance the stem *horsehair* is a **compound** of the two roots *horse* + *hair*. The stem can be modified for plural to form *horsehairs*. If we add *-er* to the root *teach* we get the stem *teacher*, to which the plural marker *-s* can be added, *teachers*. If we add *-ness* to the root *happy* we get the stem *happiness*, which can be further modified to the stem *unhappiness*, and so on.

In English roots and stems tend to be free, but there are many languages where they are bound. For instance, the verbs in the Semitic languages tend to consist of a root of only three consonants, which, through different vowel modifications, are formed into stems that then take the various inflectional markers for person and number. Thus the Hebrew root *gdr* 'to enclose' has the stem *gadar-* in the past tense, to which the person/number affixes are added: *gadar-ti* (1SG) 'I enclosed', *gadar-ta* (2SG.M) 'you (M) enclosed', and so on; but for the present tense the root is modified with a different set of vowels, to the stem *goder-*, which then inflects for person and number (Glinert 1989: 471). The root *gdr* does not function as its own word and is therefore bound. Likewise, the Spanish root *habl-* 'speak' cannot function in discourse without an ending and is therefore also bound. A stem may also be bound, for example the Columbia-Wenatchi (Salishan (Interior Salish): USA) verb *nk^{w}nám-* 'sing', which consists of a root (*k^{w}an* 'cut') and two affixes (*n-* 'in' and *-am* 'MIDDLE voice') (Willett 2003: 37). To this stem the various necessary inflectional markers are then added.

The root and stem carry **lexemic** information, i.e. the basic semantic information of the word. For instance, the lexeme of *work*, *works*, *worked* and *working* is WORK (by convention the lexeme of a word is written in small capitals to distinguish it from the various grammatical forms the words can take), the lexeme of both *hair* and *hairs* is HAIR and the lexeme of both *horsehair* and *horsehairs* is HORSEHAIR. An **affix**, on the other hand, is an obligatorily bound morpheme which does not carry any lexemic information. Affixes can be **derivational** or **inflectional**. Derivational affixes are those that create new words, for instance *un-* and *-ness* in *unhappiness* or *-ly* in *beautifully*. Inflectional affixes are those which carry grammatical information, such as *-s* 'plural' or *-ed* 'past tense'.

There are four types of affixes. A **prefix** attaches itself to the beginning of a host word. An example of a derivational prefix is *un-* in *unhappy*. Logba is an example of a language with inflectional prefixes:

Logba (Niger-Congo (Kwa): Ghana)

(18) ókpé inashína
ó-kpé i-nashína
3SG-know CM-everybody
'He knows everybody.' (Dorvlo 2008: 31)

In (18) the person/number marking and the noun class marking attaches to the beginning of the host word.

A **suffix** attaches to the end of the host word. An example of a derivational suffix is *-ness* in *happiness*. The English past tense marker *-ed* is an example of an inflectional suffix. It is far more common for languages to make use of suffixes than of prefixes in inflectional morphology. Discounting the languages in Dryer's (2011w) database that had only little affixation (141 of 971) we get 830 languages. Of these, 530 (63.9%) are suffixing[55] while 153 (18.4%) are prefixing and 147 (17.7%) are equally prefixing and suffixing.

An **infix** is an affix which places itself inside a morpheme, usually a root or a stem. For instance in Leti (Austronesian (Central Malayo-Polynesian): Indonesia), nominalizations[56] are derived from the verb through the infix *-ni-*: consider *kakri* 'to cry' > *k**ni**akri* (k-**ni**-akri) '(the) act of crying' and *pali* 'to float' > *p**ni**ali* (p-**ni**-ali) '(the) act of floating' (Blevins 1999: 400). An example of inflectional infixation can be found in Maranao (Austronesian (Southern Philippines): Philippines), where *-i-* marks past tense: *tabasan* 'slash' > *t**i**abasan* (t-**i**-abasan) 'slashed' (Reid 1992: 73 citing McKaughan 1958: 28).

We speak of a **circumfix** when at least two types of affixation have to occur at the beginning and at the end of the host at the same time. An example of an inflectional circumfix is, for example, the German past participle, which is formed by simultaneously prefixing *ge-* and suffixing *-t* to the verbal stem: *lieben* 'to love' (stem *lieb-*) > ***ge**liebt* (**ge**-lieb-**t**) '(had) loved'. Leaving out any one of the two affixes would make the construction ungrammatical; something like **gelieb* is not acceptable. An example of a derivational circumfix is the Indonesian (Austronesian (Malayic): Indonesia) *ke-… -an*, which derives (abstract) nouns, as in ***ke**bebas**an*** (**ke**-bebas-**an**) 'freedom' from the adjective *bebas* 'free' (Sneddon 1996: 35).

---

55. I have conflated Dryer's categories 'weakly suffixing' and 'strongly suffixing' as well as 'weakly prefixing' and 'strongly prefixing' into 'suffixing' and 'prefixing' respectively.

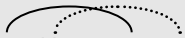
56. A nominalization is the process of forming a noun from a word belonging to some other part-of-speech, for example a verb or an adjective.

Some languages also have **parafixes**, where the two affixes that have to occur at the same time do not necessarily attach at the beginning and end of the host word. For instance, some Leti nominalizations are derived with *i-* + *-i-*, as in *natu* 'to send' > ***i**n**i**atu* (**i**-n-**i**-atu) '(the) act of sending, dispatch', *nòa* 'to advise' > ***i**n**i**òa* (**i**-n-**i**-òa) '(the) act of advising, advice' (Blevins 1999: 402). Here the combination of affixes consists of a prefix *i-* and an infix *-i-*. Another example is the Ilocano (Austronesian (Northern Philippines): Philippines) reciprocal consisting of the prefix *ag-* plus the infix *-inn-*, as in *sakit* 'hurt' > ***ag**s**inn**akit* 'hurt one another' (**ag**-s-**inn**-akit) (Schachter & Shopen 2007: 29). Just as with the circumfix mentioned above, both of these affixes must occur at the same time for the construction to be grammatical.

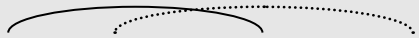
CRISS-CROSS OF AFFIXES

The Sambuǧāu dialect of Chintang (Sino-Tibetan (Bodic): Nepal) offers an interesting example of a combination of affixes.

amaikhatuptace



a-mai-kha-tup-t-a-ce



2A-[NEG]$_1$-[1DU.EXCL.P]$_1$-meet-[NEG]$_2$-PAST-[1DU.EXCL.P]$_2$

'You (SG/DU/PL) didn't meet us (DU.EXCLUSIVE).' (Bickel et al. 2007: 54)

Here we have one prefix (*a-* '2nd person agent'), one suffix (*-a* 'past tense') and two parafixes (*-mai-…-t-* 'negation' and *-kha-…-ce* '1st person dual exclusive patient'). I have glossed the parafixes in square brackets and with subscript numbers in order to better illustrate which elements belong together. Notice that both parafixes are broken up by other affixes: we do not have something like *a-mai-kha-tup-ce-t-a, where *-mai-…-t-* and *-kha-…-ce* attach symmetrically around something. Instead, *mai-…-t* is "interrupted" by the first part of *-kha-…-ce* (*-kha-…*) while *-kha-…-ce* is "interrupted" by the last part of *-mai-…-t* (*…-t-*) as well as by the past marker *-a-*, as shown by the connecting lines.

Another type of bound morpheme is **clitics**. The basic difference between clitics and affixes is that while both are phonologically dependent on a host, a clitic is syntactically independent from its host while an affix is not.[57] That is, affixes can only attach to the kinds of hosts that match their category (**part-of-speech**; for a further discussion on parts-of-speech, see Chapter 6). For instance, verbal affixes, such as the English past tense marker *-ed*, can only attach to verbs; plural marking affixes, which have to do

57. For an in-depth discussion on clitics and how they differ from affixes, see, for example, Anderson (2005).

with counting items and therefore belong to the category of nouns, can only attach to nouns, and so on. Clitics, on the other hand, are not restricted to the kind of category they may attach to. The reason they get the host that they get is because of position: they attach to the word immediately in front or after them (or inside them, as the case may be). In English, for example, the future marked *will* may cliticize to a host word. The host is always the word that immediately precedes the future marker, irrespective of what category that word belongs to. Thus the future marker *will* in the sentence *The dog will bark* may attach to *dog* to form the sentence *The dog'll bark*, even though *will* is a verbal auxiliary and belongs to the verb phrase while *dog* is a noun and belongs to the noun phrase. The reduced form *'ll* is thus a clitic and not a suffix: the reason it attaches to a word has to do with prosody and not grammar. Another difference between clitics and affixes is that clitics may function as a constituent (for more information on constituents, see Chapter 10) on their own, while an affix can never do that. Consider the Italian sentence in (19):

Italian (Indo-European (Romance): Italy)

(19) è venuto per parlarmi
è venuto per parl-ar=mi[58]
3SG.is come.PFCT to talk-INF=1SG.O
'He has come to talk to me.' (source: personal knowledge)

The infinitive affix *-are*, which gets reduced to *-ar* if a clitic attaches after it, does not form any constituent of its own but simply serves to give grammatical information about the verb. The clitic *=mi*, on the other hand, does form a constituent of its own: it is the object of the sentence.

Very often clitics are a reduced version of a free counterpart while affixes are not. For instance, the future marker in English can either be the bound form *'ll* or the free form *will*, but the past tense marked *-ed* is neither reduced, nor does it have any free counterpart.

Just as affixes may attach at different places on (or in or around) their host, so clitics may attach at different places. A **proclitic** attaches at the beginning of the host. French pronouns may attach proclitically:

French (Indo-European (Romance): France)

(20) j'attends
1SG=wait.PRES
'I'm waiting.' (source: personal knowledge)

58. Following convention, I distinguish between affixes and clitics by using a dash (-) between the affix and its host but the equal sign (=) between a clitic and its host.

A clitic that attaches at the end of the host is an **enclitic** (sometimes called **postclitic**). The Italian pronouns may attach enclitically, as shown in (19). A **mesoclitic** attaches itself between the host and the inflectional affixes. This is very rare indeed cross-linguistically, but can be found in European Portuguese[59] (Examples (21) and (22)) and in the northern dialects of Modern Greek (Example (23)).

Portuguese (Indo-European (Romance): Portugal)

(21) pedir**lhe**ia
pedir=**lhe**=ia
ask.INF=3SG.M=1SG.COND
'I would ask him.' (Vigário 2003: 270)

(22) os teus amigos emprestar**te**ão livros
os teus amigos emprestar=**te**=ão livros
3PL.M 2SG.POSS friend.PL lend.INF=2SG=3PL.FUT book.PL
'Your friends will lend you books.' (Vigário 2003: 245)

Modern Greek (Indo-European (Greek): Greece)

(23) féri**mé**ti
féri=**mé**=ti
bring.2IMP=1SG=PL
'(You.PL) bring (to) me!' (Joseph 1988: 210)

In Examples (21) to (23) above, the pronoun (bolded) cliticizes between the verbal stem and its inflectional affixes. In each of the examples, the pronoun functions as its own constituent, namely as 3rd person singular object pronoun in Example (21), 2nd person singular object pronoun in Example (22) and 1st person singular object pronoun in Example (23).[60]

Another extremely rare form of clitic is the **endoclitic**, which places itself inside the root or stem. Udi (Examples (24) and (25)) and Pashto (Example (26)) are the only two languages currently known to have endoclitics.[61]

---

59. While it may occasionally be used in Brazilian Portuguese too, this is considered extremely formal language (see e.g. Azevedo 2005). Instead, the mesoclitic is preposed to the verb, which would yield something like *lhe pediria* in Example (21) and *te emprestarão* in Example (22).

60. The Ingush (Nakh-Daghestanian (Nakh): Russia) clitic *ʔa* might also be a case of mesoclisis, since it can attach before inflectional affixes and even between the base and the reduplicated part of the base; cf. Peterson (2001).

61. I am grateful to Kathryn and Nur Khairi-Taraki for their time and effort in explaining the intricacies of Pashto pronunciation to me.

Udi (Nakh-Daghestanian (Lezgic): Azerbaijan)

(24) pasčaɣun ɣaren gölö bẹneɣsa met'alaxo
pasčaɣ-un ɣar-en gölö bẹ=ne=ɣ-sa met'a-laxo
king-GEN boy-ERG much look$_1$=3SG=look$_2$-PRES this.GEN-on
'The prince looks at this for a long time.' (root: *bẹɣ-*) (Harris 2000: 598)

(25) kaɣuzax azq'e
kaɣuz-ax a=z=q'-e
letter-DAT receive$_1$=1SG=receive$_2$-AORII
'I received the letter.' (root: *aq'-*) (Ibid.)

Pashto (Indo-European (Iranian): Afghanistan, Pakistan)

(26) a. ağustə me
wear 1SG
'I was wearing (it).'
b. á=me=ğustə
wear$_1$=1SG=wear$_2$
'I was wearing (it).' (root: *ağust-*) (Tegey 1978: 89)

In all three of the above examples, the verbal root is broken up by the pronominal clitic. The Udi verbal roots in (24) and (25) are *bẹɣ-* 'watch, look (at)' and *aq'-* 'take, receive' respectively and the Pashto verbal root in (26) is *ağust-* 'wear'. In Udi the endoclitic places itself immediately before the last element of the root. I have adopted Harris' notational convention in glossing the two parts of the root with the same translation plus a subscripted number in order to indicate which segments belong together as one lexeme. Thus the root *bẹɣ-* is chopped into the parts *bẹ-* and *-ɣ-* and the root *aq'-* is chopped into *a-* and *-q'-*. In Pashto the placement of the pronoun obeys prosodic constraints in that the pronoun immediately follows the stressed syllable of the verb: if the stress is on the last syllable of the inflected verb, the pronoun follows as a free morpheme. If, however, the stress is on the first syllable of the inflected verb, the pronoun cliticizes immediately after the stressed syllable. Again, notice that the clitic is not an inflectional affix signalling the grammar of the verb, but a completely different part-of-speech (it is a pronoun) belonging syntactically to the noun phrase and not the verb phrase. In (24) the endoclitic *ne* is part of the object, while *z* in (25) and *me* in (26) are both subjects.

5.2 Morphological typology

Since the nineteenth century and von Schlegel's classification of morphological types (cf. 1.1), languages have been classified along a linear scale of morphological typology, with isolating languages on one end and fusional languages on the other, where agglutinating languages fall in the middle; or, alternatively, adding a fourth category after

fusional languages, introflexive. This would yield something like (27) with the classic example languages given below each category.

(27)	isolating	>	agglutinative	>	fusional	(> introflexive)[62]
	Mandarin		Turkish		Latin	(Standard) Arabic

Traditionally, the view has been that whole languages could be classified in this manner, leading to statements like "Chinese is an isolating language" and "Turkish is an agglutinating language" and so on.

Bickel & Nichols (2007) have shown that this linear scale actually merges three different parameters, **fusion**, **exponence** and **flexion**, all of which are in themselves relevant for morphological typology and all of which may combine with each other. Also, any given language may employ a variety of combinations of these parameters, a fact which is obscured if we make whole-language typology statements, as already argued by Edward Sapir (see 1.1 above). In the following I will adopt Bickel & Nichol's classification of parameters. I will also adopt their approach of classifying individual morphological processes rather than whole languages.

Very simplified, the three parameters mentioned have to do with how morphemes combine. Yet another, fourth, parameter is that of synthesis, which – again very simplified – has to do with how much grammatical information a word may carry.[63] I stress once again that we are dealing with the behaviour of individual morphological processes in languages, which does not necessarily equate with the behaviour of languages as wholes. For instance, almost all languages have isolating morphemes, even if they allow other kinds of fusion. However, languages do differ as to what kinds of morphological combinations they *tend* to employ. In other words, stating that a language is *predominantly* isolating or *predominantly* non-linear may serve to give a general idea of what kinds of morphological processes to expect in the given language.

5.2.1 Fusion

Fusion denotes the degree to which morphological markers (or **formatives** in Bickel & Nichols' terminology) attach to a host stem. Following Bickel & Nichols (2007 and 2011b) I distinguish three types of fusion. A marker that stands alone as a free morpheme, that is, as an independent word, is **isolating**. Markers that are bound, i.e. that have to attach to a host, are **concatenative**. Markers that involve modifying the host in some way are **non-linear**.

---

62. This last stage essentially refers to the "root-and-pattern" typically found in Semitic languages. See below.

63. In actuality I am here merging two different concepts, that of the *phonological word* (units of form) and that of the *grammatical word* (units of grammatical analysis).

Languages may employ any and all of the types of fusion mentioned. For instance, English has isolating markers (e.g. the modal *must* in *He **must** be home by now*), concatenative markers (e.g. plural *-s* in *tree* (SG) versus *trees* (PL)) and non-linear markers (the ablaut in *s**i**ng – s**a**ng – s**u**ng*). Because of this, and because it is impossible to give a complete inventory of grammatical markers for each and every language, Bickel & Nichols (2011b) sampled their languages for grammatical case and for tense/mood/aspect (TMA) markers[64]. The patterns and tendencies mentioned below thus refer to the morphological typology of languages in these grammatical domains only. Unless otherwise mentioned, the figures in this section are based on Bickel & Nichols (2011b).

5.2.1.1 *Isolating markers*

Most languages have at least some markers that stand in phonological **isolation** and thus function as individual words. An example in English would be the modal *must*, as in *He must be in his office*. However, there are languages where all or almost all grammatical information is conveyed though isolating markers. In Koyra Chiini, for example, most markers are isolating.

Koyra Chiini (Nilo-Saharan (Songhay): Mali)

(28) ay woo kaa wor o guna
1SG.S DEM REL 2PL.S IPF see
'I here whom you (PL) see.' (Heath 1999: 97)

In (28) all grammatical information is expressed as individual words, even the tense of the verb (the imperfect marker *o*).

Isolation is not very common cross-linguistically. 16 languages (of 165 or 9.7%) in Bickel & Nichols' database are listed as exclusively isolating, mainly clustered in South East Asia and West Africa but with two Austronesian Oceanic languages (Fijian in Fiji and Rapanui on the Easter Island) and three South American languages, Wari', Kipea (Kariri (Kariri): Brazil) and Wichí (Matacoan (Matacoan): Argentina). Only one language (0.6%), Yoruba, combines isolation with non-linear (tonal) processes, while 13 (7.9%), relatively widely scattered over the world, combine isolation with concatenation.

5.2.1.2 *Concatenative (linear) markers*

The term **concatenative** literally means 'chaining together' (from Latin *con* 'with' + *catena* 'chain'). The crucial feature of concatenative markers, apart from the fact that they are bound, is that they chain together in linear strings, which means that they are segmentable. A typical example of a language with concatenative constructions is Chichewa, where the various markers attach linearly to the stems.

64. For a detailed description on how they sampled their values, see Bickel & Nichols (2011b).

Chichewa (Niger-Congo (Bantoid): Malawi)

(29)	mlenje	mmôdzi	anabwérá	ndí	míkôndo
	m-lenje	m-môdzi	a-na-bwérá	ndí	mí-kôndo
	I-hunter	I.SM-one	I.SM-PAST-come	with	IV-spears

'One hunter came with spears.' (Mchombo 1998: 518)

In (29) the grammatical markers for noun class (I *m*/*a* and IV *mi*) and past tense (*na*) are bound and are relatively straightforward to segment into morphemes.

Concatenation is a very frequent process indeed: a full 125 (75.8%) languages in Bickel & Nichols' database make use of concatenation exclusively for case and TMA marking. As mentioned above, some languages combine concatenation with isolation. Seven languages combine concatenation with non-linearity: two (1.2% of the sample) Nilo-Saharan languages (Maasai and Nandi in Kenya, both Nilotic) combine concatenation with tone while five (3%) languages, Hebrew, Egyptian Arabic (Semitic: Egypt), Middle Atlas Berber (Berber: Morocco), Beja (Beja: Sudan) (all Afro-Asiatic) and Lugbara (Nilo-Saharan (Moru-Ma'di): Uganda), combine concatenation with ablaut.

5.2.1.3 *Non-linear markers*

Non-linear markers involve some kind of modification to the host stem and are, as the term implies, not straightforward to segment into chains of morphemes. There are a number of ways that languages modify their stems nonlinearly. A very well-known strategy is found in Semitic languages, where a root consists only of a set of consonants (usually three) and where grammatical information is conveyed through insertion of a pattern of vowels, commonly termed the "**root-and-pattern**" (Ussishkin 2006: 37) but which is termed **ablaut** in Bickel & Nichols (2011b). Neither the root nor the vowel pattern can function on its own. Modern Hebrew is a language with such a pattern; an example of one kind of conjugation is the group belonging to the so-called *pa'al* verbs, as shown in (30).

Hebrew, Modern (Afro-Asiatic (Semitic): Israel)

(30) g-d-r 'enclose'

past:	a-a	(CaCaC):	gadar	'enclosed'
present:	o-e	(CoCeC):	goder	'encloses'
future:	yi-Ø-o	(yiCCoC):	yigdor	'will enclose'
imperative:	Ø-o	(CCoC):	gdor	'enclose!'
infinitive:	li-Ø-o	(liCCoC):	ligdor	'to enclose'

(Glinert 1989: 471)

In the above example, the root consonants remain the same, but the stem is modified for tense through a set of vowel combinations, none of which can be segmented into a linear string of morphemes. To indicate past tense, the root has to be modified with the vowels *a-a* to form the stem *gadar-*; to indicate present tense, the root has to be

modified with the vowels *o-e* to for the stem *goder-*; and so on. This grammatical information is not easily segmentable into a string of affixes, which makes it a non-linear process. Agreement affixes (for instance *-ti* for '1SG') may then be added concatenatively to the various stems.

Another example of ablaut (also called **gradation** or **vowel gradation**) is found in the strong verbs in Germanic languages, where inflection is marked through changes in the root vowel quality, as in English *s**i**ng* – *s**a**ng* – *s**u**ng* (present – past – past participle). Again we are not able to readily segment the words into a string of morphemes along the lines of sing-PAST or sing-PAST.PTCPL since the grammatical information is given through modifying the stem.

Ablaut is quite rare cross-linguistically; none of the languages in Bickel & Nichols' sample make exclusive use of ablaut, while only five combine ablaut with concatenation (and none combines ablaut with isolation).

Suprasegmentals (or **prosodic formatives** in Bickel & Nichols 2007), involving tone, stress and length, are another type of non-linear morphological processes. Tone is a well-known morphological strategy, common in continental South East Asia and in sub-Saharan Africa. An example of a language with grammatical tone is Lango.

Lango (Nilo-Saharan (Nilotic): Uganda)

(31) a. àpônnê
1SG.hide.PFV.MID
'I hide myself.'

b. ápònnê
1SG.hide.PROG.MID
'I am hiding myself.' (Noonan 1992: 101)

The difference between the perfective in (31a) and the progressive in (31b) is indicated through change of tone: a falling tone on *à-* plus a rising-falling tone on *-ô-* for the perfective versus a rising tone *á-* plus a falling tone *-ò-* for progressive. In neither case are the words possible to segment into a linear sequence of morphemes.

While tone is a frequent feature in the languages of the world (see 4.2.3.2), not many make use of tone to convey the grammatical information sampled in Bickel & Nichols. Six languages (3.6%) make use of tone, three of them exclusively – Iau (Lakes Plain (Lakes Plain): Indonesia), Kisi (Niger-Congo (Southern Atlantic): Guinea) and Lango – one, as mentioned above, combines tone with isolation and two, also mentioned above, combine tone with concatenation. It seems reasonable to assume that the rarity of tone in the sample is partly due to the non-proportionality to language density pointed out by Maddieson (2011f) and quoted above (4.2.3.2).

The use of stress to convey grammatical information can, for instance, be found in Italian, as in the example mentioned above (4.2.3.3): /ˈcantɔ/ 'I sing' and /canˈtɔ / 'I sang'.

An example of length serving as a morphological process can be found in the Agar Dinka (Nilo-Saharan (Nilotic): Sudan) case system, where the difference between *tǫ̱oc* 'swampy.area.ABSolutive' and *tǫ̱ooc* 'swampy.area.LOCative' (Andersen 2002: 13) is only that the locative form lengthens the vowel. In other words, there are no affixes to segment.

Replacement or **substitution** is when a regular marker replaces a part of the stem, something which is common in Nilotic languages (Bickel & Nichols 2007: 182). For instance, in Lango *-ê* is a common plural marker that attaches to the stem. However, if the stem ends in a vowel, the final stem vowel is replaced by the plural marker. Compare, for instance Example (32), where *-â* in *bórâ* is replaced by the plural *-ê*.

Lango (Nilo-Saharan (Nilotic): Uganda)

(32)	**singular**	**plural**	**gloss**	
	réc	récê	'fish'	
	bórâ	bùrê	'cat'	(Noonan 1992: 83)

Another type of replacement is **suppletion**, where a root or stem is paradigmatically replaced by a root or stem of a different etymological origin. For instance, in English the verb *to go* is inflected for past with a completely different stem, *went*, which is not a cognate (does not have the same historical origin) with *go*.[65]

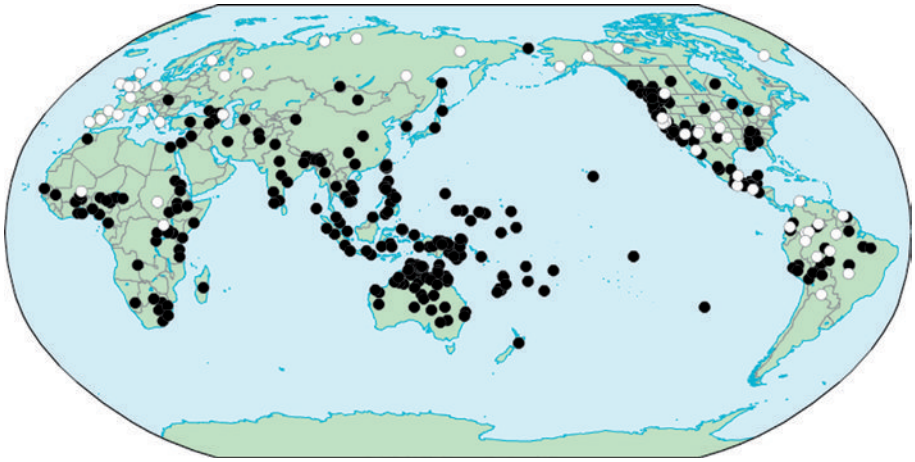
(33) She **go**es to school. (present tense) ~ She **went** to school. (past tense)

A rare type of non-linear process is **subtraction**, where the grammatical information lies in taking out an element of the stem. For instance, in Nuer (Nilo-Saharan (Nilotic): Sudan) some plurals are formed by shortening the vowel of the stem, as in *kaa̱t* 'vulture.SG' versus *ka̱t* 'vulture.PL' (Wright 1999: 33). In O'odham (Uto-Aztecan (Tepiman): USA) the perfective form is regularly derived by subtracting the final consonant from the imperfective form, as in *hi:nk* (imperfective) versus *hi:n* (perfective) 'bark' (Bickel & Nichols 2007: 183).

5.2.1.4 *Reduplication*

Reduplication falls somewhere in between concatenation and non-linear processes. Because the languages many of us are used to, such as western European languages, do not productively employ reduplication, we might not be aware of how common a strategy it is. A full 84.7% (312 of 368 languages) of Rubino's (2011) sample has productive reduplication. English is part of the minority group of 56 languages which has no productive reduplication. Map 5.1 shows the patterning of reduplicating versus non-reduplicating languages in Rubino's sample.

65. For a detailed discussion on suppletion, see Veselinova (2006).



Map 5.1 Languages with (black dots) and without (white dots) productive reduplication. Modified from (Rubino 2011). For a full legend, see http://wals.info/feature/27A.

Reduplication involves copying a set amount of phonological material from a base form (root or stem) and fusing it with that base to form a stem onto which other morphemes may then be added. It is less linear than concatenative morphemes in that the form of the reduplicant (the repeated element) is dependent on the form of the base, since it is a part of the base that is being repeated. However, it is more straightforward to segment a reduplicated form than other types of non-linear processes; once the pattern for the repetition, the reduplicative skeleton, has been identified, it can be argued that reduplication is simply "affixation … of a skeletal morpheme" (Marantz 1982: 456). The figures for the cross-linguistic patterns in this subsection are based on Rubino (2011).

Reduplication can be either **full** or **partial**, and while the reduplicant usually attaches immediately to the root it has its shape from, there are also languages with so-called **discontinuous** reduplication, where other morphological material may appear between the reduplicant and the base. Furthermore, reduplication can be **simple** or **complex**. In simple reduplication the reduplicant merely repeats a given amount of material from the base. Complex reduplication involves taking material from the base and partly altering it.

Full reduplication involves copying the whole base. Most languages allow both full and partial reduplication. However, 35 languages (9.5%) in Rubino's database allow full reduplication only. An example of a language with full simple reduplication is Erromangan, where reduplication indicates intensification.

Erromangan (Austronesian (Oceanic): Vanuatu)

(34) /unmeh/ 'early' ~ /unmehunmeh/ 'very early'
/ilar/ 'shine' ~ /ilarilar/ 'shine brightly' (Crowley 1998: 34)

An example of a full complex reduplication (also called 'echo constructions', 'echo reduplication' or 'alliterative repetition', e.g. Aikhenvald 2007: 46) can be found in Persian, where the reduplicated form changes the initial consonant to either /m/ or /p/ of the copied element. The reduplicated form takes a meaning of what we might call 'scattered generality', most closely equivalent to English 'and so forth'.

Persian (Indo-European (Iranian): Iran)

(35) *bâlâ* 'above' ~ *bâlâ**m**ala* 'somewhere above'
mive 'fruit' ~ *mive**p**ive* 'fruit and so on' (Ghaniabadi et al. 2006: 3)

Partial reduplication involves copying only a set part of the base and may involve a number of different forms. It can be a set of phonemes (C, CV, CVCV, and so on), a set of syllables or a set of morae (the minimal unit of metrical weight) that is copied (Rubino 2011). Most languages allow both partial and full reduplication, and this category, with 277 languages (75.3%), is by far the largest group in Rubino's database.

In Thao the instrumental is expressed by C*a*- reduplication, which means that the fist consonant of the base is copied and *-a-* is added (also called a **duplifix**, Haspelmath 2002: 24):

Thao (Austronesian (Paiwanic): Taiwan)

(36) *finshiq* 'to sow' ~ *fa**f**inshiq* 'seed for planting'
***f**-a-finshiq*
cput 'to filter' ~ ***c**acput* 'sieve'
***c**-a-cput*
kishkish 'to shave, cut' ~ ***k**akishkish* 'razor'
***k**-a-kishkish* (Chang 1998: 282)

Ilocano is an example of a language with several types of reduplication.

Ilocano (Austronesian (Northern Philippines): Philippines)

(37) -V- *kumrad* > *kumra**a**d* 'creak'
*kumra-**a**-d*
C- *lalaki* > *la**l**laki* 'boys'
*la-**l**-laki*
CV- *nuang* > ***nu**nuang* 'eater buffaloes'
***nu**-nuang*
CVC- *bato* > ***bat**bato* 'stones'
***bat**-bato*
CVCV- *tapiken* > ***tapi**tapikean* 'pat repeatedly'
***tapi**-tapikean*

CVCVN[66]-	*rupa*	>	***rupan****rupa*	'face to face'
			rupan*-rupa*	
	pateg	>	***patem****pateg*	'mutual caring'
			patem*-pateg*	(Rubino 2000: xvii)

An example of a partial complex reduplication can be found in Nakanai, where the vowel in the copied VC skeleton is changed to the vowel immediately preceding the reduplicant.

Nakanai (Austronesian (Oceanic): Papua New Guinea)

(38)

haro 'day' (SG)	~	*hararo* 'days' (PL)	
		*ha-**ra**-ro*	
velo 'bubbling'	~	*velelo* 'bubbling forth'	
		*ve-**le**-lo*	
hilo 'see'	~	*hililo* 'seeing' (PROG)	
		*hi-**li**-lo*	
baharu 'widow'	~	*bahararu* 'widows' (PL)	
		*baha-**ra**-ru*	

(McCarthy & Broselow 1983: 74 citing Johnston 1980: 149)

Automatic reduplication is when an affix obligatorily triggers reduplication but the reduplication itself does not add any meaning to the construction. An example of an automatic reduplication can be found in Tagalog, where adjectives are derived by prefixing *ka-*, which triggers a reduplication of the first two syllables of the word.

Tagalog (Austronesian (Meso-Philippine): Philippines)

(39)

wilih 'interested'	~	*kawilihwilih* 'interesting'	
		ka-wilih*-wilih*	
panabik 'excitement'	~	*kapanapanabik* 'exciting'	
		ka-pana*-panabik*	(French 1988: 50)

As mentioned, the reduplicant might be separated from the base by some particle. An example of such a **discontinuous** reduplication can be found in the Manila Bay Creoles, which is a cover term for Ternateño, Caviteño, and Ermiteño, where the linker *-ng-* sits between the reduplicant and the base.

Manila Bay Creoles (Creole (Spanish-lexified): Philippines)

(40)

buníta 'beautiful'	~	*bunitangbuníta* 'very beautiful'	
		bunita*-ng-*buníta	(Grant 2003: 205)

66. N stands for 'any nasal'.

In terms of proportion the languages in APiCS behave almost exactly like the languages in WALS (Michaelis et al. 2013: feature 26). After subtracting the three creoles in Rubino's sample, 84.7% (309 of 365) of the WALS languages have some kind of productive reduplication (either full or partial or both), leaving 15.3% languages with no productive reduplication. Of the languages in APiCS 86.8% (66 of 76) have some form of reduplication while 13.2% do not have any reduplication. In other words, pidgins and creoles do not seem to behave differently from non-creole languages in terms of employing the morphological process of reduplication.

5.2.2 Exponence

Languages also differ as to how many grammatical categories may be expressed by one and the same morpheme. **Separative** (or **monoexponential**) morphemes encode only one single category, while **cumulative** (or **polyexponential**, also called **portmanteau**, e.g. Booij 2005) morphemes encode several things at the same time. This parameter may interact with fusion, so that we get six logical combinations: isolating, concatenative and non-linear separative markers plus isolating, concatenative and non-linear cumulative markers. In Table 5.1 languages with examples of each of the six logical types of processes are listed.

Table 5.1 The six logical combinations of fusion and exponence.

	Isolating	**Concatenative**	**Non-linear**
Separative	Kasong (41)	Meithei (42)	Dinka (43)
Cumulative	Wari' (44)	Spanish (45)	Modern Hebrew (46)

Kasong offers an example of isolating separative markers. Each of the markers is a free morpheme, i.e. they are isolating, and each of them conveys only one piece of information, i.e. the markers are separative:

Kasong (Austro-Asiatic (Pearic): Thailand)

(41) nak kamlaŋ lɔ:ŋ ce:w prĭ
3.SG PROG FUT go forest
'S/he will be going to the forest.' (Sunee 2003: 173)

In (41) the progressive marker and the future marker both form separate words.

Meithei (Sino-Tibetan (Kuki-Chin): India) offers an example of concatenative separative markers. The markers fuse concatenatively with a host stem; they are linearly segmentable and each of the segments is separative in that each conveys only one piece of information.

Meithei (Sino-Tibetan (Kuki-Chin): India)

(42)	əynə	thəŋ	əmənə	háydu	kháy
	əy-nə	thəŋ	ə-mə-nə	háy-tu	kháy-i
	1.SG-CNTR	knife	ATT-one-INST	fruit-DDET	cut-NHYP
	'I cut the fruit with a knife.'				(Chelliah 1997: 128)

In the Dinka example mentioned above, repeated here, we have an instance of a non-linear separative process, where the absolutive and locative cases are distinguished only through phonological length. The marker (length) conveys only the information of case, and is as such separative, but it is not possible to segment from the host word, and is therefore non-linear.

Dinka (Nilo-Saharan (Nilotic): Sudan)

(43)	to̱oc	~	to̱ooc	
	'swampy.area.ABSolutive'		'swampy.area.LOCative'	(Andersen 2002: 13)

Wari' is an example a language with isolating markers, that is, the morphemes form separate words. However, they are cumulative in that they contain more than one piece of grammatical information, and this information is not possible to segment into smaller units. Example (44) thus shows an isolating cumulative process.

Wari' (Chapacura-Wanhan (Chapacura-Wanhan): Brazil)

(44)	ma'	co	tomi	na
	that.PROX.HEARER	INFL.M/FRP/P	speak	3SG.RP/P.VIC
	'Who is speaking?'			(Everett 1998: 692)

In (44) each of the morphemes are separate words, *ma'*, *co* and *na*, but each conveys a bundle of grammatical information, and none of the words can be analysed further into smaller parts. Spanish also makes use of cumulative markers that fuse concatenatively onto the stem, which gives us a concatenative cumulative morphological process.

Spanish (Indo-European (Romance): Spain)

(45)	habl-ó	
	speak-3SG.PAST.IND.PFV	
	'He spoke.'	(source: personal knowledge)

In (45) the verbal inflection marker *-ó* is bound and conveys a host of information all at the same time. In Hebrew, on the other hand, we have a non-linear cumulative process: a similar amount of information as in the Spanish example above is expressed through only one single process, but the process involves modifying the root itself and is thus non-linear. Consider, for example, the future tense of the verbal root *g-d-r* 'to enclose' mentioned above and repeated here:

Modern Hebrew (Afro-Asiatic (Semitic): Israel)

(46) g-d-r 'enclose'

future active indicative:	yigdor	'will enclose'	
future passive indicative:	yigader	'will be enclosed'	(Glinert 1989: 471)

In (46) the way the stem is modified conveys more than one piece of information: the tense, the voice, and the mood. However, this grammatical information is not linearly segmentable: if you want to change any of the grammatical information, for instance from active voice to passive voice, you have to modify the root to an entirely different stem.

It is much more common for languages to have separative morphemes than cumulative. In Bickel & Nichols' (2011a) database 127 languages (or 79.4%) have separative markers for tense/mood/aspect, while 29 (18.1%) have cumulative markers (and four are listed as not having tense/mood/aspect marking).[67] The picture differs for case marking, although again the separative markers are much more common than cumulative markers: subtracting the 75 of 162 languages that are listed as not having case (English being one of them), we are left with a total of 87, of which 71 (81.6%) have separative markers and the remaining 16 (18.4%) have cumulative markers.

5.2.3 Flexitivity

Languages also differ in how much allomorphy they have, termed **flexitivity** in Bickel & Nichols (2007). The Indo-European declension and conjugation classes are examples of flexitivity, where a set of inflectional affixes are chosen depending on which class the noun or verb belongs to. If, on the other hand, a given grammatical marker is always the same, i.e. does not vary according to classes of verbs or nouns, it is **nonflexive**. For instance, if a language has five different ways of marking the (nominative) plural, with *-e*, *-er*, *-(e)n* *-s*, or -Ø, depending on which class the noun belongs to, we have an instance of flexitivity. This is the pattern exhibited in German. If, however, the plural is always marked the same way, as is the case with Pichi (Creole (English-lexified): Equatorial Guinea) *dèn* (Yakpo 2009), we have an instance of nonflexitivity. Again it is important to keep in mind that we are dealing with individual morphological processes in languages, not making statements about the sum of the possible processes in any given language.

This is a third and separate parameter from fusion and exponence and may interact with them in various ways. The German example above is an example of flexive cumulative morphemes, because (*i*) the choice of which allomorph to take depends on which declension class the noun belongs to (flexitivity) and (*ii*) the markers express both number and case (cumulative). An example of a nonflexive cumulative marker is

67. These are Bororo (Macro-Ge (Bororo): Brazil), Maybrat (West Papuan (North Central Bird's Head): Indonesia), Sango and Tiwi (Australian (Tiwian): Australia).

the Hawai'i Creole English *wɛn* which expresses both tense (past) and aspect (perfective) at the same time (Velupillai 2003), i.e. it is cumulative. It is invariant, which makes it nonflexive. The plural marker in Pichi, however, is an example of an nonflexive separative marker because (*i*) it is invariant as the plural marker (nonflexive) and (*ii*) means only plural and nothing else (separative). An example of a flexive separative marker can be found in Warlpiri (Australian (Pama-Nyungan): Australia), where the ergative case is marked either with *-ngku* or with *-rlu* (Bickel & Nichols 2007: 185). It is flexive in that there are two alternative ways of marking ergative case, and it is separative in that it means only one thing (ergative). The four logical combinations are summarized in Table 5.2 with the languages exemplifying each type included.

Table 5.2 The four logical combinations of flexion and exponence with the languages from which representative examples are given in the text.

	Flexive	**Nonflexive**
Cumulative	German	Hawai'i Creole English
Separative	Warlpiri	Pichi

Likewise, flexitivity interacts with fusion. The German plural marking mentioned above is both flexive and concatenative; this is, in fact, the most common combination (Bickel & Nichols 2007: 186). Flexive nonlinear strategies are common in Semitic languages; we have seen that Hebrew expresses tense, mood and voice through a set of vocalisms. However, this set differs depending on which conjugation class the verb belongs to. Thus we had *a-a* for past, *yi-Ø-o* for future and *a-e* for imperative with the verbal root *g-d-r* 'to enclose' (Example (46)). For a different conjugation, such as the root *k-p-l* 'to fold', we have the vocalisms *i-e* for past (*kipel*), *ye-a-e* for future (*yekapel*) and *a-e* for imperative (*kapel*). Flexive isolating markers are very rare (Bickel & Nichols 2007: 186), but can be found in Sierra Otomí, where person and tense is marked by a free morpheme which looks different depending on what conjugation class the verb belongs to:

Sierra Otomí (Oto-Manguean (Otomian): Mexico)

(47)	**1SG.PRES**	**verb**	**conjugation class**	
	dí	pɛ̌ʔtsʔi 'I keep (it)'	I	
	dín	tófo 'I say (it)'	II	
	dídí	hóqui 'I fix (it)'	III	
	dídím	pèpfi 'I work'	IV	(Echegoyen 1979: 98ff)

The Pichi plural marking mentioned above is an example of a nonflexive isolating marker. This is pretty typical: "[n]onflexive formatives are often isolating; and the most common type of isolating formative is nonflexive" (Bickel & Nichols 2007: 187). Turkish is an example of a language where the plural marker *-lar* is nonflexive

concatenative – also a very common strategy – as it attaches to a host but is segmentable, and is invariable, i.e. is used for all nouns (Kornfilt 2003: 265). An example of a nonflexive non-linear marker is the perfective marker in Kisi, invariably expressed through a LH tone (Childs 1995: 173). The six logical combinations are summarized in Table 5.3 with the languages exemplifying each type included.

Table 5.3 The six logical combinations of flexion and fusion with the languages from which representative examples are given in the text.

	Isolating	**Concatenative**	**Non-linear**
Flexive	Sierra Otomí	German	Hebrew
Nonflexive	Pichi	Turkish	Kisi

The above sections have shown that there is much more to morphological typology than the traditional scale ranging from isolating to introflexive languages given in (27) can capture. What we have seen is that languages employ different strategies, and that these strategies themselves fall along three separate parameters that all interact with each other.

5.2.4 Synthesis

Yet another parameter is that of **synthesis**, which, very simplified, can be thought of as a scale indicating how much accumulated information a word can hold, as opposed to the parameters given above, which, again very simplified, basically denote what kinds of morphemes languages tend to have and how they combine. But bear in mind that I am simplifying matters considerably by merging the concepts of phonological word and grammatical word. For a thorough overview of word formation typology, see Aikhenvald (2007) and Dixon & Aikhenvald (2002).

There are three basic types of synthesis, which can be pictured as standing in a linear arrangement to each other:

(48) analytic > synthetic > polysynthetic

It is important here to keep in mind that this is a continuous scale and that there are no sharp boundaries between the three types. Again, bear in mind that we are dealing with morphological processes in languages, not whole language typologies. English, for instance, makes use of both analytic and synthetic constructions.

Analytic words do not take any affixation to their lexical roots or stems. An analytic way (also called **periphrastic**) of marking tense, for example, is found in the English future, as in *He will walk home*. **Synthetic** words allow affixation. An example of synthetic tense in English is the past, expressed through the *-ed* affixation, as in *He walked home*. English typically does not take a high amount of affixation. For instance,

while the grammatical coding of comparative ("more") for adjectives tends to be done synthetically if the stem is rather short, an analytic construction is favoured if the stem is rather long. Compare *strong – stong-er* or *steady – steadi-er* (synthetic) with *beautiful – more beautiful* or *helpful – more helpful* (analytic). The Chichewa Example (29) above also shows instances of synthetic words, where several pieces of grammatical information are attached to the lexical root or stem. But a synthetic word can also end up being very long. A spectacular case of synthesis can be found in Turkish:

Turkish (Altaic (Turkic): Turkey)

(49) tanıştırılamadıklarındandır
tan-ış-tır-ıl-a-ma-dık-lar-ın-dan-dır
know-RECIP-CAUS-PASS-POT-NEG-NZR-PL-3.POSS-ABL-3.COP
'It is because they cannot be introduced to each other.'
(lit. '(it) is from their not being able to be made known to each other')
(Bickel & Nichols 2007: 191)

The crucial difference between synthetic and **polysynthetic** words is that the latter involve more than one lexeme.[68] While the Turkish example in (49) is very long and involves a great deal of segments, there is only one lexeme, *tan* 'know'. Polysynthetic words, however, may contain more than one lexeme. Alutor is an example of a language with polysynthetic words. Consider the second word in (50):

Alutor (Chukotko-Kamchatkan (Northern Chukotko-Kamchatkan): Russia)

(50) gəmmə t**akka**n**nalgə**n**kuww**atavətkən
gəmmə t-**akka**-n-**nalgə**-n-**kuww**-at-avə-tk-ən
1.ABS 1SG.S-son-CAUS-skin-CAUS-dry-SUFF-SUFF-PRES-1SG.S
'I am making a son dry a skin/skins.' (Gerdts 1998: 87)

The Turkish word *tanıştırılamadıklarındandır* in (49) is as long as the Alutor word *takkannalgənkuwwatavətkən* in (50) but the Turkish word is synthetic while the Alutor word is polysynthetic. This is because the Alutor word contains three different lexemes, *akka* 'son', *nalgə* 'skin' and *kuww* 'dry' (bolded in the example). Although polysynthetic words tend to be long, they do not necessarily have to be, as (51) shows.

---

68. This is, in fact, to simplify matters a great deal. There is no one single defining feature that makes one language polysynthetic and another not; rather, languages fall on a continuum with those that are more or less synthetic cluster at one end and those that are very synthetic indeed at the other. However, typically polysynthetic languages tend to have certain features, such as, among others, a very large inventory of bound morphemes, incorporation (see 6.1.2 below), and the possibility to express entire sentences as one phonological word. See further Aikhenvald (2007: 5f).

Mamaindê (Nambikuaran (Nambikuaran): Brazil)

(51) **juk^{h}o?t^{h}**ĩntu
ju-**k^{h}o?**-t^{h}ĩn-tu
edge-hang-NCL.VILLAGE-FNS
'village hanging on the edge' (Eberhard 2009: 349)

The Mamaindê word *juk^{h}o?t^{h}ĩntu* is shorter than the Turkish word *tanıştırılamadıklarındandır*, but is still a case of polysynthesis, since it contains two lexemes, *ju* 'edge' and *k^{h}o?* 'hang' (bolded in the example).

5.3 Sign language morphology

Sign languages, just like spoken languages, have minimal meaningful units, i.e. morphemes, and instances where units may alternate, i.e. allomorphy. Morphemes may be either free, i.e. function on their own, or bound, i.e. be dependent on a host, and they may combine, for example to form compounds or derivations. In other words, signed languages are as linguistically complex as spoken languages. However, due to the fact that sign languages make use of an entirely different mode of communication, visual instead of audio, morphology in sign language tends to be less concatenative than in spoken languages (Janzen 2007). While spoken languages are dependent on the rather sequential nature of the production (and perception) of sound, signed languages have to their disposal a visual area comprising the whole upper body of the signer as well as the space around the signer. The full potential of this area is made use of, so that expressions involve not only the actual manual articulation of signs, but also various modifications. Thus grammatical and/or derivational information may be expressed simultaneously, for instance through facial gestures and spatial locations, so-called non-manual markers. In a sense that makes signed languages 3-dimensional as opposed to the 2-dimensional characteristics of spoken languages. What this amounts to is a predominantly non-linear type of morphology (Aronoff et al. 2004), although, as we shall see, sequentiality also occurs, both as affixation and as cliticization.

Compounding, which is also sequential in nature, is very common in sign languages (Sandler 2006). An example of a compound is the ASL sign for faint which consists of the signs MIND + DROP (*ibid*: 330). An example of a derivation is the ISL negative suffix, which, similar to the English *-less*, derives adjectives, for instance *shameless* in the construction SHAME + NEG[69]. This negative suffix has two allomorphs, signed either with one hand or two, depending on the host it attaches to (Meir

69. Although it originally grammaticalized from the sign NOT-EXIST, it is now a reduced and bound form of that sign. For the sake of distinguishing between the full sign NOT-EXIST and the negator, I am glossing the negator NEG. For more on grammaticalization in sign languages, see 13.3.

2004: 115ff). Examples of prefixes are the ISL 'sense' prefixes: to denote that something has to do with perception (seeing/hearing/smelling (intuiting)) a reduced and bound form of the sign for the relevant perceptory organ is prefixed to the host, for example, EYE-SHARP 'to discern by seeing' (Aronoff et al. 2003: 61).

Examples of cliticized forms occur in Turkish Sign Language (TİD: Turkey) and DGS. In TİD the negator NOT may attach itself to the preceding sign and form part of a phonological unit with that host: it (en)cliticizes. This reduced form differs from an affix in the same way as a spoken clitic differs from an affix: the clitic is not syntactically dependent on the host, but simply attaches to whatever precedes it. It also has a free counterpart (Zeshan 2004a: 46). Another example of a clitic is the DGS deictic THERE which may either occur as a free sign of form part of a single sign unit (Zeshan 2002: 166).

Non-linear morphological processes are very common in sign languages. For example, verbs are very often modified non-linearly for agreement with the subject and object or for aspect (Aronoff et al. 2003). What is non-linear about much of sign language morphology is that the base of the sign, the stem, is modified as to its rhythm, path or direction to indicate the relevant grammatical information. For instance, many verbs (so-called 'agreeing verbs' or 'directional verbs') are marked for subject and object by modifying the location-movement-location base of the stem so that location$_1$ is at the subject and location$_2$ at the object. At the same time, the palm orientation is towards the object. By having the movement component of the sign making an arch the sign has further been specified for durational aspect. All this grammatical information is expressed by taking the basic form of the stem and modifying it during the signing. In a sense this is similar to the root-and-pattern of Semitic verb forms described above, except for the fact that the Semitic root is a bound morpheme – it cannot function on its own without modification – while the sign is a free morpheme.

It seems as if sign languages universally make use of what has been termed classifiers (Aronoff et al. 2003: 63). They modify verbs and typically decode (*i*) the shape of objects, (*ii*) the handling of an object and (*iii*) the movement and location of referents. With classifiers, "the handshape of one or both hands represents a particular type of referent, while the location, arrangement and movement of the hand expresses something about the referent" (Zeshan 2002: 171). These classifiers are organized paradigmatically. An example of a complex sign using classifiers would be Figure 5.1 expressing the sentence *The car hits a tree (and gets wrecked)* in ASL. Here the non-dominant hand is configured for the classifier "tree" (the forearm upright, palm outwards) while the dominant hand is configured for "vehicle" (the hand has the thumb, index finger and middle finger stretched out while the ring finger and pinkie are bent), signs "move" (by moving the hand towards the non-dominant hand) and adds the configurations for "wrecked" at the end of the motion (index and middle finger bend).

There are two major types of classifiers, entity classifiers (encoding the referent) and handling classifiers (encoding how the referent is manoeuvred). Sign languages

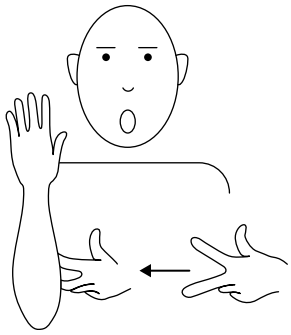


Figure 5.1 ASL verbal classifier VEHICLE (Zwitserlood 2003: 116).
Illustration: Maria Johanson. Used with permission.

vary in the amount of classifiers they have. For example, NGT has 17 classifier handshapes (Zwitserlood 2003: 138ff) while Indo-Pakistani Sign Language (IPSL: India, Pakistan) only has two, "legs" and "person" (Zeshan 2003b: 118).

Many sign languages make use of reduplication to express the general concept of "more of the same", similarly as in spoken languages. Sign reduplication is, for example, done by having the sign make an arch and thereby repeating the location-movement-location pattern in one fluid motion (which makes it an instance of a single reduplicated unit, not several repeated units). A reduplicated verb will typically indicate a longer duration of the event (durative), or that it occurs habitually (habitual), or that it occurs repeatedly (iterative). A reduplicated noun typically indicates plurality (Perniss et al. 2007: 9).

5.4 Summary

Both spoken and signed languages make use of morphemes, small units which carry information of some kind. These units can be either bound or free. The core of a lexeme is a root or a stem, the difference between the two being that the root is not further analysable into any smaller parts, while a stem may consist of a root plus something else. Affixes are bound morphemes that do not carry any lexemic information and that are syntactically dependent on what kind of host they may attach to. Clitics are also bound morphemes, but while they are phonologically dependent on a host, they are not syntactically dependent on what they may attach to. Both affixes and clitics can attach at different places on their hosts.

The traditional way of classifying languages into one of three (or possibly four) types of fusional categories obscures the fact that morphological processes make use of three different parameters. Fusion indicates how tightly morphemes attach to each other. Reduplication is a kind of fusion. Exponence indicates how much information each morpheme conveys. Flexion denotes how much allomorphy a language has. A separate, fourth, parameter is that of synthesis, which denotes how much information, both grammatical and lexemic, a word may carry.

Sign languages are as morphologically complex as spoken languages, but due to their difference in modality – spoken languages being dependent on the sequential nature of sound while signed languages have at their disposal the simultaneity of the visual medium – spoken languages are predominantly linear in their morphological processes while signed languages are predominantly non-linear.

5.5 Keywords

affix
clitic
exponence
flexion
free/bound morphemes
fusion
lexeme
morpheme
morphology
root
stem
synthesis

5.6 Exercises

1. What is the difference between an **affix** and a **clitic**?
2. Which of the words below would you call a **synthetic** one and which a **polysynthetic** one? Why?

   Mamaindê (Nambikuaran (Nambikuaran): Brazil)

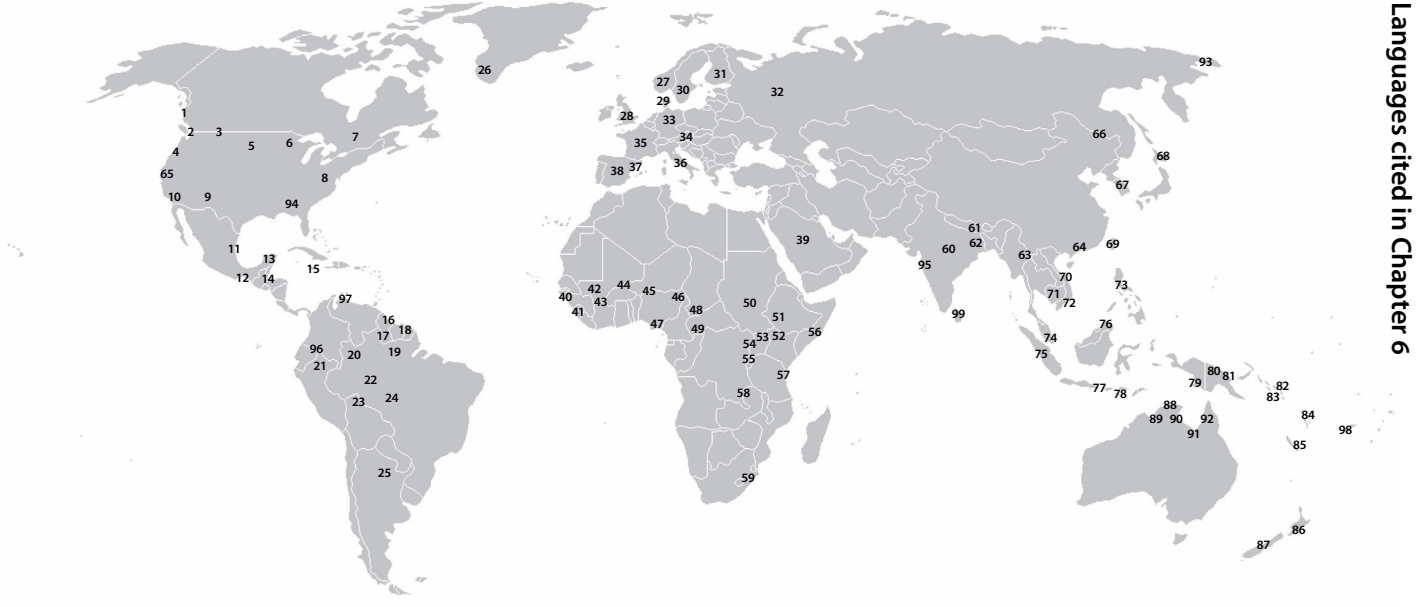
   a. jaɰ̃k-ã-oʔ-tʰũ-tu
      peccary-GEN-pound-NCL.POWDER-FNS
      'pounded meal of peccary and manioc'

   b. nũsa-jahon-nãʔã-nu-tu
      POSS1.PL-old.men-PL-FUT-FNS
      'our old men in the future'

   (Eberhard 2009: 345, 350)
3. Why is the 19th century linear scale of the morphological types isolating > agglutinative > fusional > introflexive languages inappropriate?
4. Define **reduplication** and its various forms. Discuss whether (and if so how) pidgin and creole languages differ from non-creoles with respect to this morphological process.
5. Is the following statement true or false? Motivate your answer.

   Sign languages are less morphologically complex than spoken languages.

Languages cited in Chapter 6



1 Chinuk Wawa
2 Nuuchahnulth
3 Kutenai
4 Coos, Hanis
5 Lakhota
6 Ojibwe, Minnesota
7 Quebec Sign Language
8 American Sign Language
9 Acoma
Navajo
10 Tiipay, Jamul
Maricopa
11 Nahuatl, Huasteca
12 Zoque, San Miguel Chimalapa
13 Yucatec
14 Belizean Creole
15 Jamaican
16 Berbice Dutch
17 Macushi
18 Ndyuka
Saramaccan'
19 Hixkaryana
20 Nadëb
21 Urarina
22 Pirahã
23 Wari'
24 Mamaindê
25 Mocoví
26 Greenlandic, West
27 Norwegian
28 English
29 Danish
30 Swedish
Old Norse
31 Finnish
32 Russian
33 German
German Sign Language
34 Austrian Sign Language
35 French
36 Italian
37 Catalan Sign Language
38 Spanish Sign Language
39 Arabic, Standard
40 Balanta
41 Kono
42 Bambara
43 Supyire
44 Koromfe
45 Hausa
46 Hdi
47 Igbo
48 Mbay
49 Sango
50 Krongo
51 Aari
Dime
52 Maasai
53 Lango
54 Ngiti
55 Nkore-Kiga
56 Somali
57 Swahili
58 Bemba
59 Fanakalo
60 Indo-Pakistani Sign Language
61 Limbu
62 Malto
63 Plang
64 Cantonese
65 Yana
66 Chinese Pidgin Russian
67 Korean
68 Ainu
69 Tsou
70 Pacoh
71 Khmer
72 Vietnamese
73 Chabacano, Zamboanga
Tagalog
74 Singapore Bazaar Malay
75 Javanese, Peranakan
76 Begak-Ida'an
77 Kata Kolok
78 Kambera
79 Ndom
80 Aiome
Huli
Kobon
Yimas-Arafundi Pidgin
81 Som
Takia
82 Toqabaqita
83 Lavukaleve
84 Mwotlap
85 Bislama
86 Maori
87 New Zealand Sign Language
88 Ilgar
89 Jaminjung
90 Anindilyakwa
91 Kayardild
92 Kuuk Thaayorre
93 Chukchi
94 Koasati
95 Korlai
96 Media Lengua
97 Papiamentu
98 Pidgin Hindustani
99 Sri Lankan Malay

Chapter 6

The lexicon and its classes

In all known languages, both spoken and signed, the vocabulary of an individual language can be grouped into **open** and **closed** word classes. Open word classes are typically **lexical** classes (also called **lexical categories** or **content words**) and are those where words can easily be added, for instance through derivation or other ways of forming new words, or through borrowing. Closed word classes are typically **functional** classes (also called **functional categories** or **function words**) and are those where words are not readily added; while there is change in these classes too, the change is much slower than with open classes. This chapter first gives a brief sketch of how spoken languages form new words (6.1), then moves on to give an overview of how languages tend to classify their words (6.2). I will give brief definitions of the various word class categories and show that some word classes are universal while most are not. Section 6.3 gives an overview of parts-of-speech in sign languages.

6.1 Word-formation

The following will give a brief overview of word-formation, i.e. how languages create new words for their open classes. It is by no means an exhaustive discussion of the various issues related to the topic. For very accessible overviews on word-formation, see, for example, Bauer (2003) and Booij (2005). For more detailed information, see Matthews (2006) and Aikhenvald (2007) as well as Chapters 80–93 in Booij et al. (2000).

There are two main ways for languages to form new lexemes: **derivation** and **compounding**. Compounding basically involves amalgamations of lexemes to form a new lexeme, whereas derivation relies on modifying a lexeme through various morphological processes to form a new lexeme. An example of a compound is *windmill*, formed with the lexemes WIND + MILL. An example of a derivation is *healthy*, formed with the lexeme HEALTH and the affix *-y*. Compounding and derivation are not mutually exclusive. For instance, *football* is a compound (FOOT + BALL) and from that we have the derivation *footballer* (FOOTBALL-*er*).

6.1.1 Derivation

Derivational morphology is different from inflectional morphology in that, very generally speaking, inflection carries grammatical information such as number, case and gender, while derivation does not. Derivation only serves to create new words; these new words may then take necessary inflectional morphology. For example, from *trap* we may derive the word *entrapment*, using the prefix *en-* and the suffix *-ment*. We may then inflect this derived word for plural, *entrapments* (*entrapment* + plural *-s*). Furthermore, while derivations may change the word class, for instance making a noun out of verb (such as *sing-er* from *sing*), or an adjective out of a noun (such as *health-y* from *health*), and so on, inflection never affects the word class. Also, inflection tends to be obligatory while derivation is not. For instance, in English the plural has to be marked through one of the plural allomorphs if the entities referred to consist of more than one, or the construction would be ungrammatical. However, deriving a new word using a derivational strategy is an optional choice. There are several other distinguishing features between inflection and derivation; for a thorough listing on the typical differences between the two, see Aikhenvald (2007: 36).

There are many different ways in which languages can form new words through derivation, and any one language may employ several strategies. A common derivational device is **affixing** (see 5.1 for different types of affixes). Another very common derivational device is **reduplication** (see 5.2.1.4 for examples). **Apophony** (also called **stem mutation**, Beard 1998: 62) involves internal modification of the stem, for example ablaut (see 5.2.1.3). An example of an English apophony derivation involving both a vowel and a consonant change is *breach* /bɹiːʧ/ from *break* /bɹeik/ (Aikhenvald 2007: 45). **Prosodic modification** through stress or tone is another derivational device, for instance in English, where the difference between ˈ***per****mit* (noun) and *per*ˈ***mit*** (verb) is only one of stress (see 4.2.3 for further examples).

Less common are devices which involve removing something. With **subtraction** a predictable part of the word is removed. An example can be found in French, where the masculine counterpart of the feminine adjective form is predictably shorter, namely lacking the final consonant: compare *petite* /pətit/ 'little.F' versus *petit* /pəti/ 'little.M' and *verte* /vɛʁt/ 'green.F' versus *vert* /vɛʁ/ 'green.M' (Bauer 2003: 39). Other kinds of shortenings are **truncation**, **clipping** and **back-formation**. An example of a clipping is *pram* from *perambulator* or *phone* from *telephone*. While clippings tend to mean the same thing as their longer counterparts, truncations do not, as in *evacuee* from *evacuate*. One way of thinking of truncations is that the suffix *-ate* is cut off (truncated) before the suffix *-ee* is added to *evacu-* (see e.g. Bauer 2004). Another way of seeing it is that the suffixes *-ate* and *-ee* are substituted (see e.g. Watson 1976). A back-formation is when a part of a word which seems to be an affix (but might not be) is deleted. An example is *baby-sit* from *baby-sitter* where *-er* is conceived of as a suffix parallel to the suffix in *singer* and *runner*. A **blend** involves merging two words that get partly

truncated, as in *smog* which consists of the beginning of ***sm**oke* and the end of ***fog***, or *motel* which consists of the beginning of ***mo**tor* and the end of *ho**tel***.

Conversion (also called **zero-derivation**) is when a word changes word class without any modification to the word itself. An example of a conversion would be *bottle*, which in isolation is intuitively classed as a noun, but which by conversion can be used as a verb, for instance in *To bottle wine*. Another example is *walk*, which in isolation intuitively would be classed as a verb, but which by conversion also may function as a noun, as in *To take a walk*.

6.1.2 Compounding

A very common strategy for languages to form new lexemes is through compounding. It is important to note here that a compound is not just two separate words that happen to come next to each other, but that it actually constitutes its own phonological unit. This holds true irrespective of how the compound is spelled, as it is the pronunciation that is relevant. In English, for example, compounds may either be written as one word, such as *football* (a kind of ball, also a kind of sport where that ball is used), or with a hyphen, such as *pie-eyed* (drunk), or as two separate words, such as *fire door* (a kind of door that prevents fire from spreading further); cf. also the Lango example below. The crucial thing about all these words is that they are pronounced as one phonological unit, for instance they all have only one primary stress: *football* /ˈfʊtbɔːl/, *pie-eyed* /ˈpaɪaɪd/, *fire door* /ˈfaɪədɔ(ɹ)/. There are, however, exceptions to this general rule. Pacoh (Austro-Asiatic (Katuic): Vietnam) compounds, for example, may consist of phonologically free words but still function as one single lexeme, as in *aât achéq* 'wilderness' (ANIMAL + BIRD) (Watson 1976: 226). Compounds are also typically inflected only once, as one word, according to its head (main) lexeme. If we talk about two black-and-white balls played by two opposing teams on a rectangular field with a goal at each end, we would inflect for plural only once for the entire compound and say *footballs* (not **feetballs*). Likewise, if we are talking about two fire resistant doors we would again only inflect for plural once for the entire compound, *fire doors* (not **fires doors*). Notice, again, that we are dealing with general patterns. For instance, even in English this pattern does not always hold: with the compound *tooth mark* both parts of the compound are inflected for the plural to *teeth marks*. Compounds typically do not get broken up by, for example, modifiers. We may talk about *a new football* or *a metal fire door*, but we cannot say **a footnewball* or **a fire metal door*. I stress again that while these are general tendencies, it should be kept in mind that none of these criteria are absolute universals, as the examples above have shown.

Languages may have different kinds of compounds. **Root compounds** (or **primary compounds**) typically centre on nouns one way or another. An **endocentric** compound (also called a **tatpuruṣa** compound from Sanskrit *tatpuruṣa* 'that-person') refers to "a sub-class of the items denoted by one of [the] elements" (Bauer 2003: 42).

An example of an endocentric compound in English would be *blackbird* (BLACK + BIRD), which is a type of bird. Likewise, the Maori compound *wharenui* 'meeting house' (WHARE 'house' + NUI 'big') is a type of house (Harlow 2007: 130). In both of these cases the compound is a sub-category of its head element. That is, the head element in *blackbird* is *bird* (it's a type of bird), and the head element in *wharenui* is *whare* (it's a type of house). An **exocentric** compound (also called a **bahuvrihi** compound from Sanskrit *bahuvrihi* 'having much rice'), on the other hand, is not a sub-class of any of the compounded elements. An example of an exocentric compound in English would be the above mentioned *pie-eyed*, which is neither a type of pie nor a type of eye. In Lango the compound *wàŋ ɔ̀t* 'window' (WÀŊ 'eye' + ɔ̀t 'house') is neither a type of eye nor a type of house (Noonan 1992: 115). A Maori *ihupuku* (IHU 'nose' + PUKU 'swollen') is neither a type of nose nor a type of swollen but is a sea elephant (Harlow 2007: 130). A **copulative** or **coordinate** compound (also called a **dvandva** compound from Sanskrit *dvandva* 'two-and-two') refers to "an entity made up of the two elements mentioned in the compound together" (Bauer 2003: 43). An example of a copulative compound in English would be *bitter-sweet*, denoting a taste (or feeling) which is both bitter and sweet. The word *pesa-ṭaka* 'money' in Malto (Dravidian (Northern Dravidian): India) is a copulative compound meaning the sum of its two parts *pesa* 'coin, paisa' and *ṭaka* 'bank-note, rupee' (Steever 1998: 384).

In **syntactic compounds** (sometimes also called **verbal compounds**[70]) the head element is a verb and the modifying element is something which could have functioned as that verb's argument in a phrase. An example in English would be *hair-dryer*, where the head is the verb DRY and the modifying element HAIR could have functioned as an object to the verb: *to dry hair*. An example in Russian would be *sneg-o-pad* 'snowfall' (SNEG 'snow' + *o* 'linker' + PAD 'falling'; Aikhenvald 2007: 32), where, just as in English, the head verb could have had the modifying noun as a subject: *snow falls*.

A special kind of syntactic compound is **incorporation**, which will be further discussed in Chapter 9. There is a large body of literature discussing the exact nature and properties of incorporation, most of it focussing on noun incorporation. For a very accessible overview of the major issues and theories concerned with noun incorporation, see Massam (2009). For more details, see, for example, Baker (1988), Gerdts (1998) and Aikhenvald (2007) as well as the highly influential studies by Mithun (e.g. 1984 and subsequent).[71] What makes incorporation a special type of compounding,

70. But see Aikhenvald (2007) where 'verbal compound' means root serialization. Verb serialization will be discussed in Chapter 11.

71. Notice, however, that Baker (1988) and Massam (2009) include constructions as noun incorporations that, for example Aikhenvald (2007) and Gerdts (1998) do not, such as the denominal verbs of West Greenlandic. I will essentially be following the analyses of Aikhenvald (2007) and Gerdts (1998).

and a much discussed phenomenon, is that it involves not only the word-formation process of combining two lexemes, but also involves a host of other processes, both morphological and syntactic.

> In some ways, N[oun] I[ncorporation] is the most nearly syntactic of all morphological processes. It combines constituents, namely N[oun]s and V[erbs]s, that are usually associated syntactically. It can be vastly more productive than other derivational processes, like nominalization or causativization, since it combines two potentially open sets of morphemes, N and V stems, instead of one set of stems and a limited set of affixes. (Mithun 1984: 889)

Very simplified, what happens is that the head, which is usually a verb, but can also be a preposition (Gerdts 1998: 84) absorbs a modifier, usually either a noun, pronoun or adverb, which may function as a syntactic argument (for example object) to that head. The head, however, stays in the same word class, and still functions as a verb (or preposition, as the case may be) in every respect, needing the same kind of grammatical markers, for instance for tense, person, number, and so on, as any other verb. In other words, the verb absorbs (incorporates) part of the phrase to form a complex verbal stem, which is then inflected as any other verbal stem. The incorporated element basically becomes part of the verb. This in essence has various grammatical effects, for instance with respect to syntax (especially valency) and discourse (especially information flow). These grammatical domains will be discussed in more detail in Sections 9.1.2 and 9.2.

The most common type of incorporation is a noun incorporated into a verb, **noun incorporation**. An example of noun incorporation can be found in Yucatec:

Yucatec (Mayan (Mayan): Mexico)

(52) a. t-in-p'oʔ-Ø-ah nòok
COMPL-1SG-wash-it-PERF clothes
'I washed (the) clothes.'

b. p'oʔ-nòok-n-ah-en
wash-clothes-ANTIPASS-PFV-1SG.ABS
'I clothes-washed.' (= 'I washed clothes.') (Bricker 1978: 15)

In (52a) the noun *nòok* 'clothes' stands alone as an object to the verb *p'oʔ* 'wash'. In (52b), however, that noun has been absorbed (incorporated) into the verb. Without going into too many details at this point, essentially, the result of the incorporation in Yucatec refers to "a unitary activity, in which the N[oun] modifies the type of activity predicated, but does not refer to a specific entity" (Mithun 1984: 857). That is, while in (52a) specific clothes are referred to (for example *I washed the clothes you put on the floor (but not those that were still in your bag)*) in (52b) there is no specific entity referred to, rather just a general action (for example *I washed clothes all day yesterday*).

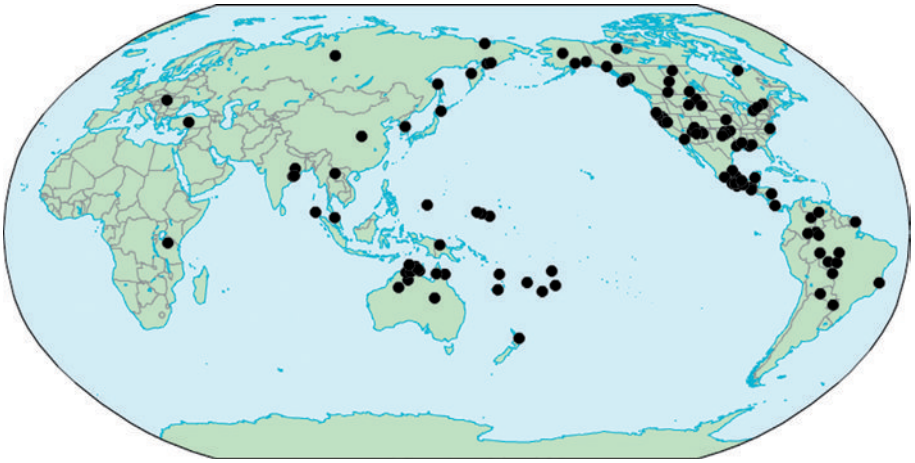
Notice that it doesn't necessarily have to be the full form of the noun that gets incorporated. Compare (53a and b) from Huasteca Nahuatl, where the incorporated noun is the stem form of the full, free noun.

Huasteca Nahuatl (Uto-Aztecan (Aztecan). Mexico)

(53) a. aškėman ti-ʔ-kʷa nakatl
never 2SG-it-eat meat
'You never eat meat.'

b. naʔ ipanima ni-naka-kʷa
1SG always 1SG-meat-eat
'I eat meat all the time.' (lit. 'I always meat-eat.') (Merlan 1976: 185)

Noun incorporation might seem a highly exotic phenomenon, but Map 6.1 shows that it is found in quite a number of languages. As can be seen from the map, noun incorporation is especially common on the American continent, both North and South.



Map 6.1 Examples of languages with noun incorporation. This is not an exhaustive survey. For a full legend, see http://dx.doi.org/10.1075/z.176.additional.

While it is rare, adpositions may also be incorporated. An example of a postposition incorporation can be found in Nadëb:

Nadëb (Nadahup (Nadahup): Brazil)

(54) a. ɛɛS a-hɨng hxɔɔh go
father FORMATIVE[72]-go.downriver canoe in
'Father goes downriver in a canoe.'

72. In Nadëb the verbal root is bound and needs some prefix to form a stem. The 'formative prefix' is used to form a verbal stem if there are no other prefixes attached to the verbal root.

b. hxɔɔh$_O$ ɛɛ$_A$ ga-hɨng
canoe father in-go.downriver
'Father goes downriver in a canoe.' (lit. 'Father goes-downriver-in (a) canoe.')
(Martins & Martins 1999: 262)

The postposition *go* 'in' in (54) has been incorporated into the verb *hɨng* 'go-downriver' in (54) to form *gahɨng* 'go.downriver.in' (I am ignoring the assimilation process of *go* to *ga* here). The effect here is that the object (*hxɔɔh* 'canoe') gets emphasized.[73]

Because of the difficulty in distinguishing it from cliticization, it is beyond the scope of this book to discuss pronoun incorporation. Basically, a clitic simply attaches to its host, while pronoun incorporation creates a new lexeme. A kind of example that might serve as an illustration is the difference in English between *bathe* and *bask*, where the latter verb (*bask*) is the result of a fusion of the Old Norse (Indo-European (Germanic): present-day Scandinavia) reflexive pronoun *sik* 'self' with the verb *baða* 'bathe' which led to the new lexeme BASK.

Chukchi (Chukotko-Kamchatkan (Northern Chukotko-Kamchatkan): Russia) offers an example of how adjectives may get incorporated into nouns, as in *nilgəqin quoraŋə* '(a) white reindeer' (ABS.SG) versus *elgəquorata* '(by a) white.reindeer' (ERG.SG) (Muravyova 1998: 527).

6.2 Parts-of-speech

In the previous section, as well as previous chapters, such terms as *noun*, *verb*, *adjective*, *adverb*, *adposition* and *pronoun* have occurred in abundance. But what exactly do they mean? Well, they are different **parts-of-speech** (or **word classes**), which in essence are major categories of words that group together grammatically. This section will give a brief and very simplified overview of the four lexical class parts-of-speech that may occur in a language, as well as some typical functional class parts-of-speech found in the languages of the world. For very concise overviews on word classes, see Evans (2000), Haspelmath (2001) and Anward (2006). For more details and discussion, see, for example, Givón (2001a: 49ff) and Schachter & Shopen (2007). For a somewhat different approach to lexical classes, see, for example, Baker (2004), Rijkhoff (2007), Hengeveld & van Lier (2010). It would be beyond the scope of this section to give an exhaustive list of all the functional parts-of-speech we know about; for more details, see Schachter & Shopen (2007), which this section relies heavily on.

73. In actuality it is a case of manipulating the valency of the verb, which then affects which word(s) will function as obligatory objects. This in turn influences certain discourse properties, such as what element gets focussed on, and so on. See further Section 9.2.

Languages differ radically in how many classes they have and in the proportions of these classes. Some languages have an extremely limited set of closed class words (or functional categories), while others have a high number of such words. Some languages have only two open word classes (or lexical categories), others, like English, have as many as four separate such classes. Furthermore, a word class found in one language will not necessarily be found in another language. In other words, while it seems to be universal that languages actually do group their words into categories of some kind, the categories themselves are language independent.

6.2.1 Lexical classes

As mentioned above, it seems that all known languages distinguish between open and closed classes. The open classes typically consist of **content words**, i.e. words with more or less concrete, specific meanings. Languages may have up to four major open class parts-of-speech, *nouns*, *verbs*, *adjectives* and *adverbs*. The definitions of these categories rely on a cluster of features, both semantic (denoting meaning), grammatical and syntactic (how items are combined).

Each of the above-mentioned categories may contain further subcategories. In English, for example, nouns can be subcategorized into mass and count nouns, depending on whether they can take the plural (e.g. *sand*/**sands* versus *chair*/*chairs*); or proper and common nouns, depending on whether they can take the article (e.g. *Peter*/**the Peter* versus *chair*/*the chair*), or abstract versus concrete (e.g. *emotion* versus *chair*), and so on. Other languages subcategorize depending on whether or not the item is possessable. In Maasai, for example, nouns are either possessable or non-possessable. Such things as tools, money, houses, kin, and so on can be marked for possession grammatically, but such things as meat, water, land and stars cannot (or if they are marked for possession it sounds very odd to the speakers). So while it is acceptable to mark *ɛnkɛ́rái* 'child' for possession, *ɛnkɛ́ráy áy* 'my child', it is not readily acceptable to mark *ɛnkʊ́p* 'land, dirt' for possession, ??*ɛnkʊ́p áy* 'my land' (Payne 1997: 40). Yet other languages, for instance Mamaindê, subcategorize their nouns depending on, among other things, physical properties such as consistency (whether the item in question is solid or liquid) and shape (Eberhard 2009). In other words, the potential subcategorizations of each major part-of-speech category are language dependent.

It is important to keep in mind that the defining characteristics given for each part-of-speech category should not be seen as absolute, but as general indications of features that typically cluster together in a given word class.

6.2.1.1 *Nouns*

The first major open part-of-speech category, **noun**, typically refers to things, persons and places, but also includes abstract notions such as feelings, ideas and so on. Grammatically, nouns may typically be marked for number (how many of the item(s) are being referred to), case (what role the item has in the sentence), gender (what subcategory the item belongs to) and definiteness (whether it is a specific entity referred to or not), for instance through morphological processes, but also, especially in the case of languages with predominantly analytic strategies, through syntactic processes. Nouns may also combine with demonstrative pronouns (e.g. *this/that* as in *this/that house*) and may function as arguments (that is, participants, e.g. subject and object) in a clause. For an article-length overview of the noun, see Lehmann & Moravcsik (2000).

English has two numbers, singular (one entity) and plural (more than one of the same entity), as in *chair* versus *chairs*. Other languages, such as Lavukaleve, specify for dual (two of the same entity), as in *funfun* 'firefly' (singular) – *funfunil* '(two) fireflies' (dual) – *funfunaul* 'fireflies' (plural) (Terrill 1999: 97). For the number values trial (three of the same entity) and paucal (a few of the same entity) see 7.1.1.1.

While English hardly has any case marking at all, the exception being the genitive *'s* as in *chair's*, many languages do mark for case. An example of a language that marks its nouns for case is Dime (Afro-Asiatic (South Omotic): Ethiopia): compare *zití* 'ox' (nominative case) with *zitím* 'ox' (accusative case) (Seyoum 2008). For more on case systems, see 7.1.3.

Gender refers to which subclass the noun belongs to. In French, for example, nouns are either masculine (***le*** *cadeau* 'the gift') or feminine (***la*** *table* 'the table'), while in German they are either masculine (***der*** *Stuhl* 'the chair'), feminine (***die*** *Mütze* 'the hat, cap') or neuter (***das*** *Buch* 'the book'). Swedish also has two genders, but unlike French they are not masculine or feminine, but neuter (***ett*** *träd* 'a tree') and "non-neuter" or common (***en*** *bok* 'a book'). See further 7.1.2.

Definiteness indicates whether we are referring to a general example of an entity or a specific entity, as in the difference between ***a*** *man* and ***the*** *man*.

6.2.1.2 *Verbs*

The second major open part-of-speech category, **verb**, typically refers to actions and processes (e.g. *dance*, *grow*, etc.), but also states (e.g. *know*, *exist*, etc.).[74] Grammatically, verbs may typically be marked for tense (placing the event in time), aspect (specifying

74. It is not a universal that verbs belong to an open class of words: there are languages that have a distinct but closed class of verbs, found predominantly in Australia and Papua New Guinea. Jaminjung (Australian (Jaminjungan): Australia) is an example of such a language, with a class of verbs containing only 33 members (Schultze-Berndt 2000: 84).

the perspective taken on the event), mood (indicating the speaker's attitude toward a situation or a statement), voice (e.g. whether an event is active or passive) and various devices for manipulating valency (specifying how many obligatory arguments the verb must have), as well as polarity (whether the statement is in the affirmative or the negative). Again, as with nouns, these operations may be marked either through morphological or syntactic processes. Verbs may also be marked for person agreement, where a grammatical marker indicates the number (one or more) and person (first, second or third) of an argument, most commonly the subject. Verbs typically function as predicates, typically form the core of the sentence or clause and typically have "a relational meaning, relating one or more participants (or arguments) to an event" (Anward 2006: 408). For an article-length overview of the verb, see Bybee (2000).

English has three tenses, two of which are marked morphologically and one that is marked syntactically. The present tense (placing the event in the present, the 'now') is marked with a suffix *-s* for third person singular, as in *He walk**s***. The past tense (placing the event in the past) is also marked morphologically, most commonly with a suffixed *-ed* (for the moment I will ignore other means of marking the past tense in English), as in *He walk**ed***. The future tense (placing the event in the future) is marked analytically with the use of an auxiliary verb (see below), as in *He **will** walk*.[75] Other languages have other kinds of tense systems, such as only two tenses (either past versus nonpast or future versus nonfuture), or have more than three, or none. For more on tense, see 8.2.

It is common for languages to have some kind of aspect marking. English, for example, marks for progressive (denoting that the event is on-going) with the suffix *-ing*, as in *He is walk**ing***. Languages may also make a grammatical difference between perfective (extremely simplified, if an event is seen as a completed whole) and imperfective (again extremely simplified, if an event is seen as an ongoing process). French is an example of such a language, where the difference between *Il a payé* 'he paid' (perfective) and *Il payait* 'he paid' (imperfective) is one of aspect. See further Section 8.4.

English typically expresses modality other than indicative (the form typically used for declarative sentences) analytically, except for the imperative (giving commands), which is the base form of the verb, as in *Walk!* Languages may mark for wishes (desideratives and optatives), for prohibitions (prohibitive), for tentativeness or hypothetical situations (subjunctive), for what kind of evidence they have for a statement (evidentials), and so on. For more on mood and modality, see 8.5.

---

75. This is of course again a simplification, as there are several ways of indicating futurity in Modern English. For example, the progressive *–ing* form may indicate futurity: *He is coming tomorrow*.

English has two voices, active and passive. An example of an active sentence is *He opened the door*, which, in the passive, would be *The door was opened by him*. The difference between the two is one of valency, which will be further discussed in 9.2.

Languages also mark for polarity one way or another, and often that is done in connection with the verb. English contrasts affirmative and negative sentences with *not*, as in *He walked* versus *He did not walk* (I am, for the moment, ignoring the obligatory auxiliary verb for English negations). Other languages mark negation through an affix on the verb. In Dime, for example, the negative is marked by *-káy*: *nú ʔadéén* 'he comes' (affirmative) versus *nú ʔadkay* 'he does not come' (negative) (Seyoum 2009: 190). For a discussion on the position of the negative morpheme relative to the verb surveyed for 1326 languages, see Dryer (2011m).

The only form of person agreement that English has is the present tense third person singular suffix *-s*, as in *He walk**s*** versus *I walk*. Other languages, however, grammatically indicate agreement for all three persons, and all the numbers that the language has. In Italian, for example, the verb is inflected for three persons and two numbers:

Italian (Indo-European: (Romance): Italy)

(55)				
	1SG	mangio	'I eat'	
	2SG	mangi	'you eat'	
	3SG	mangia	'he/she/it eats'	
	1PL	mangiamo	'we eat'	
	2PL	mangiate	'you eat'	
	3PL	mangiano	'they eat'	(source: personal knowledge)

Other languages may inflect for dual, and even trial and paucal, depending on their systems. See further Section 9.1.3.5.

I stress again that none of these criteria are absolutes, but rather form a cluster of characteristics that may serve to identify whether a given word is a noun or a verb. For example, there are languages, such as Mwotlap, where nouns may take tense, mood and aspect markers.

Mwotlap (Austronesian (Oceanic): Vanuatu)

(56) kōyō ma-tayak kē, tō kē ni-ēntē-yō togolgol
3DU PFCT-adopt 3SG then 3SG AOR-child-3DU straight
'They have adopted him, so that he (became) their legitimate son.'
(François 2005: 131)

In (56) *ēntē* 'child' is marked both for tense/aspect with the AORist prefix *ni-* and for agreement with the 3rd person DUal subject (they-2, the two of them) with the suffix *-yō*, just as if it had been a verb. In Nuuchahnulth the same word may translate either as a noun phrase or a verb phrase:

Nuuchahnulth (Wakashan (Southern Wakashan): Canada)

(57) inikw-ihl-minh-ʻis-it
fire/burn-house.LOC-PL-small-PAST
'The little fires that were once burning in the house.' OR
'Several small fires were burning in the house.' (Baker 2001: 25f)

In (57) *inikw* 'fire/burn' may translate either as a noun (*fire*) or as a verb (*burn*), depending on context. In other words, languages might allow a significant amount of overlap between the characteristics of nouns and verbs. However, while the difference between the two classes might not be as clear-cut as in English, there is still a distinction between them. For instance, in Nuuchahnulth verb stems (or verbal predicates, to be more precise) can only function in the same contexts as nominals when modified by the enclitic article =ˀ*iq*/=ʔ*iˑ* (M/N) (Davidson 2002: 91ff, 324ff). In Mwotlap the tense, mood and aspect markers are obligatory for verbs, while they are only optional for nouns (François 2005). One might think of it as a continuum on a scale between the two ends in Figure 6.1, where on the one end of the scale there is very little overlap between the characteristics of nouns and verbs, such as in English, and on the other end there is a lot of – but not complete – overlap between the characteristics of nouns and verbs, such as in Nuuchahnulth or Mwotlap. It thus seems that a distinction at some level between two open classes, nouns and verbs, is a near-absolute universal.

Figure 6.1 A continuum of overlap between characteristics of nouns and verbs. Languages with very little amount of overlap, such as in English, fall on the left end of the scale, while languages with a lot of overlap, such as Nuuchahnulth, fall on the right end of the scale. Notice, however, that even at the right-most edge of the scale there is not complete overlap between the two.

6.2.1.3 *Adjectives*

The third group of words which may constitute an open class, **adjectives**, typically denotes qualities or attributes, such as colour, size, shape and so on. Adjectives typically modify nouns. Notice that this section will deal with descriptive adjectives only and not those noun modifiers usually termed either quantitative or limiting adjectives, such as *many*, *some*, *a few* and so on. These kinds of noun modifiers never constitute an open class. Descriptive adjectives, however, may form an open class in many languages, though this is by far not universal. Grammatically adjectives may be specified

for degree (the extent to which a property holds), either morphologically or syntactically, and may combine with degree words that cannot combine with nouns or verbs. An example of the latter in English is *too*: while it is possible to say that something is *too cold*, constructions like **too table* or **too run* are not acceptable. In some languages adjectives show agreement in form with the noun they modify. In German, for example, the adjective is marked morphologically to agree with the gender of the noun it modifies: *ein rot**er** Stuhl* (masculine) 'a red chair' versus *eine rot**e** Blume* (feminine) 'a red flower' versus *ein rot**es** Haus* (neuter) 'a red house'. For an article-length overview of the adjective, see Bhat & Pustet (2000).

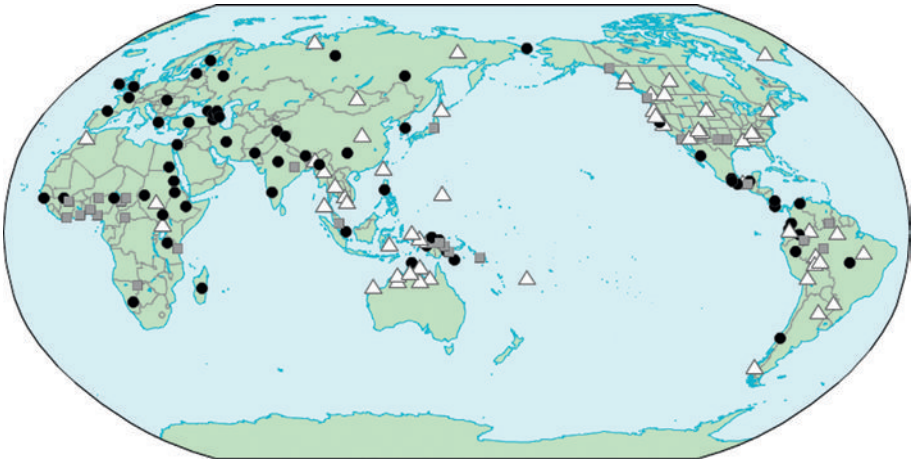
The traditional degree modifications make a distinction between three degrees: positive, comparative and superlative. As mentioned, degree or comparison may be expressed either morphologically or syntactically. English is a language with both options, where longer adjectives typically take an analytic degree modification. For instance, the comparison *tall* (positive) versus *tall**er*** (comparative) versus *tall**est*** (superlative) is marked morphologically, while the comparison *beautiful* (positive) versus ***more** beautiful* (comparative) versus ***most** beautiful* (superlative) is marked analytically.

The use of adjectives (or the equivalent) falls into two groups: that of modification of a noun or that of predication (denoting a property of the subject of a clause). An example of the former is *a big apple*, while an example of the latter is *The apple is big*.

While nouns and verbs form near-universal open class categories, this is not the case with adjectives. For those of us who are used to European languages, this might seem exotic. However, it is not all that rare for languages to either have a closed class of adjectives, or to not have a distinct word class for adjectives at all. In a pilot survey I mapped 153 languages for adjectives, with three values: 'open class', 'closed class' and 'no separate class' (Map 6.2). The language sample is based on the WALS 200-sample,[76] but it should be noted that at this stage it is not entirely balanced. For instance the number of Indo-European languages is proportionally too high in comparison to, for example, Niger-Congo or Austronesian languages. Nevertheless, as a pilot survey it serves as a starting point and does give some indications as to possible geographic patterns. Of the 153 languages, 66 (43.1%) have an open class for adjectives, while 30 (19.6%) have a closed class and 57 (37.3%) do not have any separate class for adjectives. As for genealogical patterns, all of the Indo-European languages sampled have an open adjective class, all but two (Balanta (Northern Atlantic: Senegal) and Bambara) of the Niger-Congo languages have a closed class of adjectives, and all but one of the Australian languages, Ilgar (Iwaidjan: Australia), lack a distinct adjective class.

---

76. While not every language overlaps with those in the 200-sample, every genus does. That is, if I did not have access to a specific language, I chose another language from the same genus.



Map 6.2 Adjectives in a pilot study sample. Black dots: open class (66 languages); grey squares: closed class (30 languages); white triangles: no separate class (57 languages). For a full legend, see http://dx.doi.org/10.1075/z.176.additional.

Geographically we do seem to get some patterns. For instance, a large continuous area spanning all of Europe, the Middle and Near East, as well as most of South Asia show black dots. Mainland East Asia (both South and North) as well as Northern Australia show mainly white triangles. West Africa predominantly shows grey dots. North and South America are predominantly covered in grey or white, but notice that Meso-America almost only displays black dots.

A language with a very small closed class of adjectives is Igbo, counting only eight of them, as shown in Table 6.1.

Table 6.1 Igbo (Niger-Congo (Igboid): Nigeria) adjectives.

VALUE		COLOUR		DIMENSION		AGE	
ọ́má	'good'	ójíʔí	'black, dark'	úkwú	'large'	ọ́hṵ́ʔrṵ́	'new'
ọ́jọ́ʔọ́	'bad'	ọ́cá	'white, light'	ńtá	'small'	ócyè	'old'

(Welmers & Welmers 1969: 321)

The adjectives of Igbo form a neat example of Dixon's (1982) findings that the four properties 'dimension', 'colour', 'age' and 'value' are those most likely to be found in a closed class of adjectives, while other properties, such as position (*high*, *low*), human character (*kind*, *evil*), speed (*fast*, *slow*) and physical characteristics (*hard*, *soft*) are more likely to be expressed with either nouns or verbs in languages with a closed class

of adjectives.[77] An example of a language where such properties are expressed by nouns is Hausa, where the literal translation is 'having X':

Hausa (Afro-Asiatic (West Chadic): Nigeria)

(58) a. mutum mai alheri
person having kindness
'a kind person'
b. mutum mai doki
person having horse
'a person having a horse'
c. yana da alheri
he.is with kindness
'He is kind.'
d. yana da doki
he.is with horse
'He has a horse.' (Schachter & Shopen 2007: 15)

In Example (58) shows how descriptive properties such as kindness (*alheri*) are expressed by possession of nouns. The constructions are identical to expressing other kinds of possession. Thus the attributive quality in Example (58a) is constructed in the same way as the possessive in (58b), and the predicative quality in Example (58c) is expressed in the same way as the possessive in (58d). An example of a language that expresses adjectival notions with verbs is Bemba, where the verb is either relativized (59a) or not (59c), depending on whether it is a structure modifying a noun or whether it is a predicative clause.

Bemba (Niger-Congo (Bantoid): DR Congo)

(59) a. umuuntu ùashipa
person who.is.brave
'a brave person'
b. umuuntu ùalemba
person who.is.writing
'a person who is writing'
c. umuuntu áashipa
person is.brave
'The person is brave.'

77. An example of a language with a very small closed class of adjectives indeed is Toqabaqita (Austronesian (Oceanic): Solomon Islands), which has only one dimension adjective *kali*/*kasi* (with the plural form *kaala*) 'small'. All other noun modifications are expressed with stative verbs or nouns. See further Lichtenberk (2005).

(59) d. umuuntu áalemba
person is.writing
'The person is writing.' (Schachter & Shopen 2007:16)

In other words, while all languages have ways of describing things, not all have a special class of words for that task, as Map 6.2 shows.[78]

6.2.1.4 *Adverbs*

The fourth and last group of words that may form an open class, **adverbs**, is arguably the most heterogeneous of all word classes. Basically and very simplified, adverbs typically modify categories other than nouns. Thus adverbs may modify verbs (*run **quickly***), adjectives (***quite** happy*), other adverbs (***very quickly***), prepositions (***right** out*, ***well** within*), and so on, but not nouns (**dog quickly*, **quite dog*, **right dog*, **well dog*). Notice, however, that adverbs may modify noun phrases (constructions where the noun is the main component), as in *That was **quite** [a party]*. Because the adverbs form such a mixed group, five main subclasses are usually distinguished: (*i*) **setting adverbs** of space and time (*here*, *there*, *below*, *above*; *now*, *then*, *today*, *never*); (*ii*) **manner adverbs** (also called **predicate adverbs**; *quickly*, *repeatedly*, *well*, *badly*); (*iii*) **degree adverbs** (*very*, *too*, *extremely*); (*iv*) **linking adverbs** (also called **text adverbs**; *however*, *therefore*, *hence*, *thus*); (*v*) **sentence adverbs** (*unfortunately*, *probably*, *maybe*, *frankly*). For a very detailed and accessible overview of adverbs and their characteristics, see Quirk et al. (1985). Setting, degree and linking adverbs typically form closed subclasses within the class of adverbs even in those languages where adverbs form an open class. Sentence adverbs are generally rare and seem to be a characteristic of the written languages of Europe (Haspelmath 2001). This means that the only subclass of adverbs that is actually open is manner adverbs.

It is quite common for languages to form manner adverbs from adjectives. This is the case in English, where adverbs are easily derived by adding *-ly* to an adjective: *slow* > *slow**ly***, *beautiful* > *beautiful**ly***, *happy* > *happi**ly***, and so on. As with adjectives, languages differ in whether adverbs form an open, a closed, or no class at all. In

78. But see Dixon (2006a), where he puts forth the argument that just as there may be significant but not complete overlap between nouns and verbs, the same holds for adjectives. He postulates that even those adjectives that form subclasses of verbs or nouns actually do differ in subtle ways from other verbs or nouns in those classes. In many languages this is indeed the case. For instance in Lango properties are denoted by verbs. However, adjectival verbs differ from other verbs in three respects: (i) in tone, (ii) in mood, and (iii) in the fact that some adjectival verbs have different stems for singular and plural (Noonan 1992:104). In Thai, which also lacks a special class for adjectives, only those verbs functioning as adjectives may appear in comparative and superlative constructions (Iwasaki & Ingkaphirom 2009:10).

Modern Standard Arabic, for example, adverbs form a closed class and most adverbial meanings are conveyed by adjectives or nouns in the accusative case (Ryding 2005). Thus *ɣadan* 'tomorrow' is the accusative form of the noun *ɣadu* 'next.day' and *sariʕan* 'swiftly' is the accusative form of the adjective *sariʕ* 'swift' (Schachter & Shopen 2007: 21).

There are also languages without any separate class for manner adverbs. In Swedish, for example, manner adverb expressions are expressed with the adjective in the neuter form:

Swedish (Indo-European (Germanic): Sweden)

(60) a. tåget är långsamt
train.DEF.NEUT is slow.NEUT
'The train is slow.'

b. han läser långsamt
he reads low.NEUT
'He reads slowly.' (source: personal knowledge)

In Example (60) the adjective *långsam* 'slow' is inflected in the neuter, *långsam**t***, to make it agree with the neuter noun *tåg* 'train' which it modifies. In (60b) the same word, i.e. the neuter form of the adjective, functions as a manner adverb, modifying the verb *läsa* 'read'.

Ainu has neither a special class for adjectives nor a special class for adverbs; in both cases the stative verb is used:

Ainu (Isolate: Japan)

(61) a. pirka menoko
good woman
'pretty woman'

b. pirka inu
good listen
'listen well'

c. tunasno pirka
quick good
'Get well quickly!' (Shibatani 1990: 80)

In Example (61a) the stative verb *pirka* 'be.good' is used as an adjective, modifying the noun *menoko* 'woman'. In (61b), however, the same word is used as an adverb, modifying the verb *inu* 'listen'. A stative verb may also modify another stative verb, as shown in (61c), where the optional reading of 'become X' allows for the imperative reading which would literally translate into something like "Become good fast!".

Due to the amount of overlap we have seen between various lexical classes, Hengeveld et al. (2004) propose the following implicational hierarchy (given here in a slightly modified version):[79]

(62) Verbs > Nouns > Adjectives > (Manner) Adverbs

What this hierarchy implies is that the further to the left a category is, the more likely it is that it exists as its own lexical class in a language. It also implies that if a language has a separate open class for nouns, then it also has a separate category for verbs, and if a language has a separate open class for adjectives, then it also has separate open classes for nouns and verbs. This means that the hierarchy postulates that a language with a separate open category for adverbs necessarily has a separate open class for adjectives. However, this hierarchy only illustrate tendencies and not universals: in Jaminjung, for example, verbs form a closed class of words (cf. footnote 74 above) and in Hixkaryana, adjectives have been analysed as actually belonging to the category of adverbs (Meira & Gildea 2009).

6.2.2 Functional classes

Closed word classes typically consist of **function words** (or **grammatical words**), i.e. words with abstract, general meaning, or even no meaning at all but merely a grammatical function.[80] While open classes tend to have many members – the reason they are called 'open' is because new words are readily added – closed classes tend to be small in size, since what makes them closed is the fact that new words are not readily added. As mentioned, it seems a near-universal for languages to differentiate between open and closed classes. In other words, most known languages have at least two open classes (nouns and verbs) and one or a few closed classes. Languages differ considerably not only in the number but also in what types of closed classes they have. There seems to be a correlation between the kind of synthesis a language tends to allow and the

---

79. The original terminology is:

Head of predicate phrase	> Head of referential phrase	> Modifier of referential phrase	> Modifier of predicate phrase

(Hengeveld et al. 2004: 533)

Notice, however, that this hierarchy specifically refers to syntactic slots and does not take into account other grammatical differentiations. Notice also that 'Modifier of predicate phrase' refers to manner adverbs exclusively.

80. Although there are languages where content words form a closed class, such as verbs (cf. footnote 74 above).

amount of closed word classes it has: languages with predominantly analytic constructions employ function words to a higher degree than languages with predominantly synthetic constructions (Schachter & Shopen 2007). This is hardly surprising, since the grammatical information expressed by function words in predominantly analytic languages is expressed through affixation in predominantly synthetic languages.

LIMITED CLOSED CLASSES

Yana, an extinct Hokan language that was spoken in the USA, had a very meagre inventory of closed class items: a small set of articles, a few interjections and a proclitic case marker. That was all. (See Schachter & Shopen 2007: 23f with references).

This section will bring up some reasonably common functional class categories, but makes no attempt to be exhaustive in the list of known closed classes. For a detailed overview of most of the known closed classes, see Schachter & Shopen (2007).

6.2.2.1 *Pronouns*

Pronouns are used to substitute a noun or a noun phrase.[81] This is usually a large and diverse group of closed class words. Languages differ as to what types of pronouns they have, or if they have any at all. Commonly recognized subtypes are personal, possessive, demonstrative, indefinite, relative, reflexive, reciprocal and interrogative pronouns. None of these subcategories are absolute but vary depending on language as well as on theoretical orientation. Languages may employ affixes or clitics to express the various pronominal functions. For a detailed discussion on pronouns from a cross-linguistic perspective, see Bhat (2004). For an article-length overview, see Schwartz (2000).

Personal pronouns typically refer to the speaker(s) (*I*, *we*), the addressee(s) (*you*), as well as other things that the context makes clear (*s/he*, *it*, *they*). Many languages distinguish between three persons (1, 2, 3) and two or three numbers (either singular and plural or singular, dual and plural). Common personal pronoun paradigms are (I am, for the moment, ignoring various politeness distinctions):

---

81. For an overview of other kinds of pro-forms, such as pro-sentences, pro-clauses, pro-verbs, pro-adjectives, pro-adverbs (replacing sentences, clauses, verbs, adjectives and adverbs respectively) and so-called interrogative pro-forms (e.g. the English *who*, *what*, *where*, *when*, also called content question words), see, for example, Schachter & Shopen (2007: 31ff).

German (Indo-European (Germanic): Germany) (source: personal knowledge)

	SINGULAR		PLURAL	
1	ich	I	wir	we
2	du	you	ihr	you
3	er, sie, es	s/he, it	sie	they

Plang (Austro-Asiatic (Palaung-Khmuic): China) (Suchada 2004: 57)[82]

	SINGULAR		DUAL		PLURAL	
1	ʔuiʔ^{R2}	I	ʔiiʔ^{R1}	we (two)	ləʔuiʔ^{R2}	we
2	miiʔ^{R1}	you	piiʔ^{R1}	you (two)	lemiiʔ^{R1}	you
3	ʔənR1, ʔaiʔ^{R1}	s/he, it	ləʔənR1	they (two)	ləʔənR1, ləʔaiʔ^{R1}	they

Begak-Ida'an (Austronesian (Northwest Malayo-Polynesian): Malaysia) (Goudswaard 2005: 94)

	SINGULAR		PLURAL	
1-I			kito	I + you (PL)
1-E			kəmmi	I + others
1	aku	I		
2	ikow	you	muyu	you
3	rumo	s/he, it	(m)iro	they

Kuuk Thaayorre (Australian (Pama-Nyungan): Australia) (Gaby 2006: 86ff)

	SINGULAR		DUAL		PLURAL	
1-I			ngal	I + you (SG)	ngamp	I + you (PL)
1-E			ngali	I + other	ngancn	I + others
1	ngay	I				
2	nhunt	you	nhip	you two	nhurr	you
3	nhul	s/he, it	pul	they two	peln	they

In the first paradigm (German) we find a system common to the languages of Europe, where three persons and two numbers are distinguished: *I*, *you* (SG), *he*/*she*/*it*, *we*, *you* (PL), *they*. In the second paradigm (Plang), a third number distinction is added, that of dual, denoting two of the same. The third and fourth paradigms (Begak-Ida'an and Kuuk Thaayorre respectively) show systems with an inclusive (1-I) versus exclusive (1-E) distinction. **Inclusive** means that the addressee is included while **exclusive** means that the addressee is not included in the group referred to. This might seem difficult to grasp for those of us who are not used to such systems; one way of thinking about

82. The superscript codes quality of voice: R1 means that the word is pronounced with a normal voice while R2 means that the word is pronounced with a breathy voice.

it might be to picture a religious person addressing God and asking for forgiveness. If, when praying for "forgiveness for our sins", the inclusive pronoun would be used, then the addressee, i.e. God in this case, would be included in the group of sinners! In such a situation it is highly likely that a speaker of a language with inclusive/exclusive distinctions would use the exclusive form, where "us" refers to 'I and others' but not the addressee.[83] Making some kind of inclusive/exclusive distinction is actually not as exotic as one might think if one is mostly used to European languages. About one third (68 or 34%) of the 200 languages in Cysouw's (2011) sample differentiate between 'we.INCL' and 'we.EXCL' (none of them pidgins or creoles).[84] While the figures for this feature are at the time of writing still temporary for the APiCS languages, they indicate that only about an eighth have an inclusive/exclusive differentiation (Michaelis et al. 2013: feature 15). Pidgin and creole languages thus seem less likely than non-creoles to make this differentiation.

Languages may also have paradigms where only person but not number is coded. In his sample of 261 languages, Daniel (2011) found 9 (3.4%) languages that do not differentiate formally for number in their independent personal pronoun system. An example of such a language is Mamaindê where *tại* simply means 'first person' i.e. may translate into either 'I' or 'we' (or any other number, e.g. dual), *wại* means 'second person' ('you' singular or plural), and *hãi* means 'third person' ('he/she/it/they') (Eberhard 2009: 375). English constitutes a mixed type of language: the first and third persons are distinguished for number, while the second is not. Roughly 20% of the languages in Daniel's sample have a mixed system; most of them differentiate in number in the second ('you.SG' versus 'you.PL') and third persons ('he/she/it' versus 'they'), as opposed to English, which differentiates in the first ('I' versus 'we') and third persons (Daniel 2011). While it is rare, there are also languages that do not have independent personal pronouns, such as Mbay (Nilo-Saharan (Bongo-Bagirmi): Chad), where the function of independent pronouns is carried out by affixes on other word classes, such as verbs, nouns or adpositions (Keegan 1997). Only two languages (0.8%) in Daniel's (2011) sample lack independent personal pronouns: Acoma (Keresan (Keresan): USA) and Wari'.

Demonstrative pronouns serve to point out something in a situation or a sentence. Examples in English are *this*/*these* and *that*/*those*. Demonstratives typically indicate distance between what is being referred to and the speaker (or, as the case may be, the hearer). For instance, in English *this* implies a certain closeness to the

83. Frederic Baraga, Slovene bishop and documenter of Minnesota Ojibwe (Algic (Algonquian): USA), indeed did make the mistake of using the inclusive 'we' and thus including God in the group of sinners in his Ojibwe liturgical translations (Anthony Grant, p.c.).

84. This conflates Cysouw's (2011) values 'Only inclusive' and 'Inclusive/exclusive'.

speaker while *that* implies a certain distance to the speaker. The relative differentiation is, however, fuzzy. Thus it is quite acceptable for a customer to point at an item very near to him/her and say 'I'll take that too'. It is common for languages to make two or three distance distinctions, but there are some with only one (that is, where there is no distance contrast) and some with as many as four distinctions. An example of a language where the demonstrative pronouns do not have any distance distinctions is Koromfe (distance distinctions are expressed with adjectives). There is one neutral demonstrative pronoun, *nɛŋ* 'thus':

Koromfe (Niger-Congo (Gur): Burkina Faso, Mali)

(63) nɛŋ la gʊ mɔ̄ sogɑ
thus COP PRON.3SG.NHUM also boundary
'This is the end of it.' (i.e. the story) (Rennison 1997: 258)[85]

English is an example of a language with two distance differentiations in its demonstrative pronoun system: *this* (proximal or near) *that* (distal or further away). Hdi is an example of a language with a three-way contrast of distance: *ná* 'proximate', *yá* 'middle distance' and *á* 'remote'.

Hdi (Afro-Asiatic (Biu-Mandara): Cameroon)

(64) a. bà-f-b-í tá ná-ná[86]
build-UP-build-SG OBJ DEM-DEM
'I build this.'
b. bà-f-b-í tá yá-yá
build-UP-build-SG OBJ DEM-DEM
'I build that.'
c. bà-f-b-í tá á-á
build-UP-build-SG OBJ DEM-DEM
'I build that (over there).' (Frajzyngier & Shay 2002: 84f)

Kambera (Austronesian (Central Malayo-Polynesian): Indonesia) is an example of a language with a four-way distance contrast: *ni* 'near/at speaker', *nai* 'middle distance from speaker', *na* 'near addressee' and *nu* 'far from both speaker and addressee' (Klamer 1998: 55f). Diessel (2011) surveyed languages for distance distinctions in adnominal demonstratives, i.e. those that are not independent but must occur with the noun they refer to, as in French ***ce** livre-**la*** 'that book' as in *donne-moi ce livre-la* 'give me that book' (lit. give-me this book-there); something like **donne-moi ce*, without the noun, would

85. Determiners may be used as demonstrative pronouns in Koromfe. The choice of determiner/demonstrative is based on the animacy and number of the noun it refers to (Rennison 1997: 259).

86. The demonstrative in Hdi is reduplicated when used as an independent pronoun.

not be possible. The WALS figures are thus not for pronominal demonstratives (i.e. those which replace a noun or noun phrase), yet they are interesting; Diessel found that the most common strategy (with 127 of 243 languages or 54.3%) was for languages to have a two-way contrast and the second most common strategy to have a three-way contrast (88 languages or 37.6%). Only eight languages (3.3%) made use of a four-way contrast and four[87] (1.6%) of a five-way contrast, while seven languages (3%) lacked any distance contrast altogether. The figures are at the time of writing still temporary for this feature in APiCS (Michaelis et al. 2013: feature 33), but they seem to indicate that almost two thirds of the APiCS languages have a two-way contrast while almost one fifth lack any distance contrast. This would imply that pidgin and creole languages are more prone to lack distance contrasts and less prone to have three-way contrasts than non-creole languages.

Reflexive pronouns typically denote an entity which is identical with another grammatical argument (usually the subject) in the same clause. That is, they are co-referential with a co-occurring nominal. In English this is expressed by adding *-self/-selves* to a form of the personal pronoun: *myself*, *yourself*, *himself*, *herself*, *itself*, *ourselves*, *yourselves*, *themselves*. In many languages, English among them, the same form may also be used as an emphatic, as in *He himself did it*. A little over half of the languages sampled by König & Siemund (2011), 94 (of 168 or 56%), have identical forms for the emphatic and the reflexive. The remaining 74 (44%) differentiate between the two. An example of a language that differentiates between the reflexive and the emphatic is German. Compare:

German (Indo-European (Germanic): Germany)

(65) a. Das Mädchen sah **sich** im Spiegel
ART.N girl see.PAST REFL PREP.DAT mirror
'The girl saw herself in the mirror.'

b. Der President **selbst** hielt eine Rede
ART.M president EMPH hold.PAST ART.F talk
'The president himself gave a talk.' (source: personal knowledge)

The pattern differs somewhat in APiCS, where the two strategies each comprise about one third of the languages: in 27 of 69 languages[88] (39.1.3%) the two forms are identical, while in 25 (36.2%) they are differentiated (Michaelis et al. 2013: feature 88). The remaining quater consists of languages that either have a mixed system or lack reflexive

87. These are Koasati (Muskogean (Muskogean): USA), Malagasy (Austronesian (Barito): Madagascar), Maricopa (Hokan (Yuman): USA) and Navajo.

88. Information is missing for some of the APiCS languages, bringing down the total to 69 from 76.

pronouns altogether. Languages that do not have specific free forms for reflexive meanings may employ verbal affixes to express reflexivity.

Reciprocal pronouns are also co-referential with a co-occurring nominal, but the crucial thing about reciprocals is that they also express mutuality, as in the English *each other* and *one another* (e.g. *They supported each other*/*one another*). Again, languages differ with respect to reciprocal constructions. More than half of the languages in Maslova & Nedjalkov's (2011) sample have two formally distinct forms for reciprocity and reflexivity (99 of 175 or 56.6%), as English does. In about a quarter of the sample (44 languages or 25.1%) they are formally identical, as in Wari':

Wari' (Chapacura-Wanhan (Chapacura-Wanhan): Brazil)

(66) wac xucucun hwijima'
cut REFL/RECIP.3PM children
'The children cut themselves. / 'The children cut each other.' (Everett 1998: 186)

In Example (66) the same form, *xucucun*, may translate either into 'themselves' (reflexive) or 'each other' (reciprocal). A minority of languages (16 or 9.1%) have a mixed system, where the reflexive marker may be used as a reciprocal but where there is also a separate reciprocal form, as in German:

German (Indo-European (Germanic): Germany)

(67) a. sie mögen sich
3PL like.3PL REFL/RECIP
'They like themselves.' / 'They like each other.'
b. sie mögen einander
3PL like.3PL RECIP
'They like each other.' (source: personal knowledge)

While *sich* in (67a) is ambiguous and can have either a reflexive or a reciprocal meaning, *einander* in (67b) is unambiguous and can only have a reciprocal meaning. The remaining 16 languages (9.1%) in Maslova & Nedjalkov's (2011) sample do not have any formal ways of marking reciprocal, but instead employ very iconic constructions such as repeating the verb or clause. An example of such a language is Cantonese:

Cantonese (Sino-Tibetan (Chinese): China)

(68) léih hóyíh bōng ngóh ngóh hóyíh bōng léih
2SG can help 1SG 1SG can help 2SG
'We can help each other.' (lit. 'You can help me I can help you.')
(Matthews & Yip 1994: 87)

The languages in APiCS show a slightly different picture (Michaelis et al. 2013: feature 89). A slightly higher proportion (47 of 69 or 68.1%)[89] than the languages in the WALS sample make a formal distinction between reciprocity and reflexivity. The proportion of mixed-system languages is also slightly higher in the APiCS sample than the WALS sample, with nine (13%) languages. The proportion of languages having identical forms, however, is slightly lower, with ten (14.5%) languages. Finally, the proportion of languages without any reciprocal construction at all is lower, with only three (4.3%) languages: Belizean Creole, Bislama (Creole (English-lexified): Vanuatu) and Fanakalo (Pidgin (Zulu-lexified): South Africa).

Languages that do not have reciprocal pronouns but have other kinds of reciprocal markers tend to indicate reciprocity through verbal affixes.

Relative pronouns serve to introduce a modifying clause within the noun phrase. The English relative pronouns are *who*, *whom* and *which*.

(69) The man *who* phoned yesterday.
The girl *whom* I talked to.
The chair *which* I sat on.

In Example (69) the relative pronouns *who* and *which* introduce a clause (*X phoned yesterday*/*I talked to X*/*I sat on X*) which serves as a modifying element to the noun phrase (*the man*/*the girl*/*the chair*). What a relative clause does, very simplified, is to delineate a specific antecedent (element referred to) to which a certain proposition is true. In other words, *The man who phoned yesterday* points out that particular man (out of a potential of several different men) who phoned the day before (and, for instance, not the one who phoned today), and *The chair which I sat on* points out the particular chair (out of several potential chairs) which I sat on (and not, for example, the empty one). Languages have different strategies for expressing relativity, and a given language may combine strategies. English, for instance, makes use of relative pronouns, but also allows a gap strategy as in *The girl Ø I talked to*, where there is no relative marker at all. In fact, while the gap strategy is very common cross-linguistically, comprising 125 of 166 languages (or 75.3%) in Comrie & Kuteva's (2011b) sample, to make use of a relative pronoun is actually quite rare, found in only 12 (or 7.2%) languages in the sample. The proportions differ minimally if the contact languages are subtracted, with 123 of 164 languages (75%) having the gap strategy and 12 (7.4%) having a relative pronoun. While the figures for the languages in APiCS are at the time of writing still temporary for this feature, they indicate that these languages differ from the above mentioned pattern. The APiCS languages seem to have a higher proportion of languages with a relative pronoun (about one fifth) and a much lower proportion of languages making

89. Information is missing for some languages, which is why the total is 69 instead of the usual 76.

use of a gap strategy (just under two fifths) (Michaelis et al. 2013: feature 92). For more details on relative clauses and various relativization strategies, see 11.2.3.

Indefinite pronouns differ from the pronouns discussed above in that they are used to refer to non-specific entities. In English this is expressed by such pronouns as *somebody*, *someone*, *something*, *anybody*, *anyone*, *anything*. To have such transparent pronouns as English (*some*/*any* + *body*/*one*/*thing*), based on generic nouns, is not as common as those of us who are very used to English might think: 85 of 326 (26.1%) languages in Haspelmath's (2011b) sample make use of this strategy. It is especially common in Africa, Papua New Guinea and on the Pacific islands. Most commonly (194 languages or 59.5%) indefinite pronouns are based on interrogatives (see below), as in Russian *ktoto* 'someone' (from *kto* 'who' + *-to* 'INDEF') and *čtoto* 'something' (from *čto* 'what' + *-to* 'INDEF') (Haspelmath 2011b). Rather rare is to have either special forms, unrelated to anything else, found in only 22 (6.7%) languages of the sample, or to have a mixed system (23 languages or 7.1%), where the two forms behave differently. An example of a language with a special indefinite form is Swedish with the special *någon* 'someone' and *något* 'something', historically derived from an entire Proto-Scandinavian clause: **ne wait ek hwariʀ* (*hwariʝaʀ*) 'I don't know who' (lit. 'not know I who') whence also the Danish *nogen* and the Norwegian *noen* (Hellquist 1980: *sv*). An example of a language with a mixed type of interrogative is Khmer (Austro-Asiatic (Khmer): Cambodia), where *kè:* 'someone' is based on the generic noun 'person' while *ʔvɣy*(*-mú:əy*) 'something' is based on the interrogative *ʔvɣy* 'what' (Haspelmath 2011b citing Jacob 1968). Extremely rare is to not have any indefinites at all, but to express the equivalent meaning through an existential construction. There are only two such languages (0.6%) in Haspelmath's sample, Mocoví (Guaicuruan (Guaicuruan): Argentina) and Tagalog (Example (70)).

Tagalog (Austronesian (Meso-Philippine): Philippines)

(70) may d<un>arating araw-araw
EXIST <ACTOR VOICE>come.IPFCT day~RED
'Someone comes every day.' (lit. 'There exists (one who) comes every day.')
(Schachter & Otanes 1983: 276)

The pattern for the languages in the APiCS differs radically from the pattern in Haspelmath's sample. The latter contains only one contact language (Sango) which affects the figures minimally. With the APiCS languages the most common strategy by far is to have a generic noun-based indefinite pronoun (Michaelis et al. 2013: feature 21). This is found in a 50% of the sample (38 languages). Only three languages (3.9%) have interrogative-based indefinites: Chinuk Wawa (Pidgin (Chinook-lexified): Canada, US), Singapore Bazaar Malay (Pidgin (Malay-lexified): Singapore) and Sri Lankan Malay (Creole (Malay-lexified): Sri Lanka). The proportion of languages with

a special form is roughly the same among the APiCS languages as in Haspelmath's sample (5 languages or 6.6%),[90] while the proportion of languages with a mixed system is higher (10 languages or 13.2%). Only one language (1.3%) expresses indefinite meaning through an existential construction: Zamboanga. 17 APiCS languages (22.4%) have the value 'Other'. Data is missing for Media Lengua and Yimas-Arafundi Pidgin.

Interrogative pronouns are pronouns used to form content questions. English interrogative pronouns are *who* and *what*. Interrogatives and the kinds of strategies languages have for forming content questions will be discussed further in Section 12.1.2.

6.2.2.2 *Adpositions*

Adpositions are words that express the relation between the noun phrase they govern and some other element in the clause or sentence.[91] They may come either before, after or even within the noun phrase they govern and typically express temporal, spatial or similar connections.

Prepositions precede the noun phrase they govern. This is common cross-linguistically: in Dryer's (2011i) very large sample, 512 of 1185 (43.2%) languages have prepositions. English is one of them, with such prepositions as *in*/*near* (***in*** *the house*), *to*/*from* (***to*** *the sea*), *on*/*under*/*beside*/*above* (***on*** *the table*), *by* (***by*** *the side*).

Postpositions follow the noun phrase they govern. Languages with postpositions are somewhat more common than those with prepositions in Dryer's database, comprising 48.7% (577 languages). An example of a language with postpositions is Aari. Example (71) shows how the postpositions *zan* 'top, superior location' and *dar* 'to' follow the noun phrases they govern.

Aari (Afro-Asiatic (South Omotic): Ethiopia)

(71) a. diirá **zan** dóqse
grass on sat.3SG
'He sat on (the) grass.'
b. gában **dar** káyye
market to went.3SG
'He went to (the) market.' (Hayward 1990: 489f)

---

90. The five languages are Jamaican (Creole (English-lexified): Jamaica), Korlai (Creole (Portuguese-lexified): India), Pidgin Hindustani (Pidgin (Fiji Hindi-lexified): Fiji), Saramaccan (Creole (English-lexified): Suriname) and Papiamentu (Creole (Spanish-lexified): Netherlands Antilles).

91. They govern the noun phrases in that they not only link to them, but also typically require that the noun phrase they link to be modified grammatically one way or another. For instance *to* governs the oblique case: *to him* (but not **to he*). For more on government and agreement, see 7.2.2.

Inpositions, which occur inside the noun phrase they govern, form a very rare type of adposition. Only eight languages (0.7%) in Dryer's sample have this type of adposition. It should be noted that Dryer here includes clitics, such as in the example below.

Anindilyakwa (Australian (Anindilyakwa): Australia)

(72) namwirntakakpwarthanaka [akini=**lhangwa** apwirtha]
3:1PL.were.scared.INTS 3:4.that=ABL 3:4.whale
'They were all very scared of the whale.' (lit. …the-of whale) (Leeding 1989: 312)

In Example (72) the clitic *lhangwa* 'about' is placed inside the noun phrase *akini apwirtha* 'that whale'.[92]

Some languages (58 or 4.9%) have more than one type of adposition. The languages in this group usually have both prepositions and postpositions, though there are also cases of languages with both postpositions and inpositions, such as Hanis Coos (Oregon Coast (Coosan): USA) (Dryer 2011i). There are also languages that do not have adpositions at all, such as Kutenai (Isolate: Canada). This group comprises 30 (2.5%) languages of Dryer's sample.

The figures for the languages in APiCS are at the time of writing still temporary for this feature, but it seems that they differ radically from the pattern in Dryer's sample, where the figures change minimally once Ndyuka (Creole (English-lexified): Suriname) and Sango have been subtracted (both prepositional languages). To have prepositions as the dominant strategy seems much more common in the APiCS sample, occurring in something like seven eighths of the languages (Michaelis et al. 2013: feature 4). Only very few seem to have postpositions as their dominant strategy. This seems to suggest that it is more likely that a pidgin or creole language will have prepositions than it is that a non-creole will have prepositions.

6.2.2.3 *Numerals*

Numerals are sets of words used to indicate the precise number of something. They typically have characteristics of both open and closed class words. **Cardinal numerals** are used to express the number of individuals in a set, as in *four cars*. **Ordinal numerals** are used to express the rank in a series, as in *the fourth car*. Grammatically numerals may either constitute their own closed class, or overlap with other word classes in the language. In Finnish (Uralic (Finnic): Finland), for instance, the numeral inflects for case just like nouns (Sulkala & Karjalainen 1992) while in Krongo (Kadugli (Kadugli): Sudan) numerals behave like verbs (Reh 1985). In English numerals form their own class.

92. This is in the source treated as a 'peripheral case' clitic but conforms to Dryer's definition of adpositions in that "it combines with a noun phrase and indicates the grammatical or semantic relationship of that noun phrase to the verb in the clause" (Dryer 2011i).

In most languages larger numbers are built by combining smaller numbers, or bases. English, for example, has 10 as its numeral base, called a decimal system: the numbers 1–10 are expressed by unique words, while higher numbers are expressed by multiples of 10 (e.g. *six-teen* '6 + 10' or *thir-ty-one* '3 × 10 + 1'). This is a very common strategy, found in 125 of 196 (or 63.8%) languages in Comrie's (2011c) sample. A vigesimal system, not uncommon, uses a base of 20, found in 42 (21.4%) languages in Comrie's sample. This conflates Comrie's categories 'pure vigesimal' and 'hybrid vigesimal-decimal' (where the system is vigesimal up to 100 but then switches to decimal, as is the case in Danish). A base-20 system as defined in Comrie (2011c) is found in Chimalapa Zoque (Mixe-Zoque (Mixe-Zoque): Mexico), as in *ʔiʔpšaŋʔ makkanh* '30' (20 + 10) and *tuhtaŋ ʔiʔpšaŋʔ* '120' (6x20) (Johnson 2000: 414f). Other number bases are base-2, found in Aiome (Lower Sepik-Ramu (Annaberg): Papua New Guinea) with *nogom* '1', *omngar* '2', *omngar nogom* '3' (2 + 1) (Harrison 2007: 188); base-3, found in Som (Trans-New Guinea (Finisterre-Huon): Papua New Guinea) (Hammarström 2010: 9); base-4, found in Ngiti (Nilo-Saharan (Lendu): DR Congo) (Kutsch Lojenga 1994: 357); base-5, found in Supyire (Niger-Congo (Gur): Mali) (Carlson 1994: 167); base-6, found in Ndom (Kolopom (Kolopom): Papua New Guinea) (Harrison 2007: 191); base-12, found in northern Nigeria (Hammarström 2010: 12ff) and base-15, found in Huli (Trans New Guinea (Engan): Papua New Guinea) (Cheetham 1978: 16). Languages may also lack a numeral base, in that there are no numerals above one, found in Amazonian languages (Hammarström 2010: 8).

That which Comrie calls a restricted system, found in 20 (10.2%) languages of his sample, denotes "a numeral system that does not effectively go above around twenty" (Comrie 2011c). An example of this is Kayardild with a number system going up to four but not further: *warirra* 'nothing', *warngiida* '1', *kiyarrngka* '2', *burldamurra* '3', *mirndinda* '4' and *muthaa* 'many' (Evans 1995: 242). A language with a very restricted numeral system indeed is Pirahã, which has no numerals at all (Everett 2005: 526). Some languages use body parts to extend their numeral system. This is found in only 4 (2%) languages in Comrie's sample, and is a feature known to be concentrated in Highland New Guinea, although it is also found in other parts of the world (Harrison 2007: 174). An example of such a language is Kobon (Trans-New Guinea (Madang): Papua New Guinea), which functions as follows:

> to count from 1 to 12: [on the left side of the body] little finger, ring finger, middle finger, index finger, thumb, wrist, middle of forearm, inside of elbow, middle of upper arm, shoulder, collarbone, hole above breastbone. The count can then continue down the right-hand side of the body, from the collarbone to the (right) shoulder as 13 to the little finger as 23. It is then possible to reverse the count, starting from the little finger of the right hand as 24 back up to the hole above the breastbone as 35 and down again to the little finger of the left hand as 46. (Comrie 2011c)

6.2.2.4 *Articles*

Articles are used to indicate whether the noun phrase referred to is identifiable or not. This is typically expressed in terms of definiteness: a definite item is presumed to be possible to identify (***the*** *book*) whereas an indefinite item is not assumed to be identifiable (***a*** *book*). For an article-length overview of articles, see Schwartz (2000).

Definite articles encode specificity and are used with an identifiable noun or noun phrase, for example *the book* in *Give me* ***the*** *book*, where the use of the definite article *the* indicates that the speaker assumes that the addressee is able to identify the specific book referred to. A definite article may precede the noun phrase, as in English, or follow it, as in Lakhota.

Lakhota (Siouan (Siouan): USA)

(73) chā' ki hā'ske
tree the tall
'The tree is tall.' (Van Valin 1977: 36)

In Example (73) the definite article *ki* 'the' follows the noun (*chā*' 'tree') it specifies.

Languages may use the demonstrative to indicate definiteness. In his sample of 620 languages, Dryer (2011a) found 216 (34.8%) languages where the definite article was distinct from the demonstrative (as the case is in English) and 69 (11.1%) languages where the demonstrative is used to indicate definiteness. An example of the latter type is Takia, where the demonstrative (*y*)*en*/*an*/*on* 'this (particular)' (*e*/*a*/*o* if followed by one or more modifiers) is used to mark definiteness, as shown in (74):

Takia (Austronesian (Oceanic): Papua New Guinea)

(74) yu o inug an sa-n biouŋ
war DEM formerly DEM POSS-3SG clothes
'The things from the first war.' (Ross 2002b: 224)

Some languages (92 or 14.8%) do mark definiteness, but through affixation and not separate words, as in the Swedish construction *boken* 'the book' (*bok-**en*** 'book-DEF'). Others (45 or 7.3%) do not have definite articles, but do have indefinite ones (see below). The second biggest group in Dryer's sample (198 languages or 31.9%), however, has neither definite nor indefinite articles. Korean is an example of such a language. The reading of Example (75) is thus ambiguous with respect to definiteness.

Korean (Isolate: N, S Korea)

(75) ai ka pang ey ki-e tul-e w-ass-ta
child NM room to crawl-INF enter-INF come-PST-DC
'The/A child crawled into the/a room.' (Sohn 2001: 267)

The above figures change insignificantly when the two contact languages Ndyuka and Sango, both having a definite word distinct from the demonstrative, are subtracted. The languages in the APiCS display a different pattern (Michaelis et al. 2013: feature 28). While here too the biggest group (38 languages or 50%) distinguishes between the definite article and the demonstrative, the second largest group, with 20 languages (36.3%), is actually the one where the two forms are identical. Nine languages (11.8%) do not have any definite article but do have indefinite ones, while nine (11.8%) have neither type. None of the APiCS languages make use of affixation to indicate definiteness.

Indefinite articles are used when an entity referred to is not presupposed to be identifiable by the addressee, for example *a pencil* in *Give me **a** pencil*, where no specific pencil is referred to. While we might be used to the fact that 'a' (indefinite article) and 'one' (numeral) are distinct words, it is actually equally common to use the same word for the two expressions. In his sample of 534 languages, Dryer (2011e) found 102 (19.1%) languages that behave like English, i.e. where the indefinite article and the numeral are two distinct forms (*a book* versus *one book*). In roughly the same proportion of languages, 112 (21%), the same form is used, as in Swedish *en bok* 'a/one book'. There are thus two ways of translating a sentence like (76) into English.[93]

Swedish (Indo-European (Germanic): Sweden)

(76) jag har en bok
1SG have ART/NUM book
'I have a book.'/'I have one book.' (source: personal knowledge)

In 24 (4.5%) languages indefiniteness is marked through affixation. An example of such a language is Limbu (Sino-Tibetan (Bodic): Nepal), as in the construction *yaŋdhik* 'a rupee' (*yaŋ-**dhik*** lit. 'money-a') (van Driem 1987: 32). However, the biggest group by far in this sample is the group of languages that neither have any definite nor any indefinite article (see Example (75) above). These figures differ minimally once Ndyuka (where the indefinite word is the same as 'one'), the only contact language in the sample, has been taken out.

Again the pattern exhibited in the APiCS sample differs (Michaelis et al. 2013: feature 29). About a quarter (20 or 26.3%) of the languages differentiate between the indefinite article and the word for 'one', while more than half (46 languages or 60.5%) have identical forms. One (1.3%) language, Yimas-Arafundi Pidgin, lacks indefinite articles altogether but has definite articles, and, as mentioned above, nine languages (11.8%) have neither definite nor indefinite articles. This seems to indicate that pidgin

93. This is actually somewhat simplified, as stress serves to disambiguate the two: if *en* is stressed, it means 'one' and if it is unstressed it means 'a'.

and creole languages are more likely than non-creoles to use the same form for indefinite article 'a' and the numeral 'one'.

6.2.2.5 *Auxiliaries*

Auxiliaries are semantically more or less empty verbs conveying mainly grammatical information. While some auxiliary verbs can also be used as full verbs (e.g. *have*) and thus have a semantic content, when they function as auxiliaries their primary task is not to convey meaning but grammatical information. A verb phrase containing auxiliaries will thus also contain a **lexical verb** or 'main' verb, which carries the semantic content of the construction. For example, the verb phrase *will jump* contains the auxiliary *will* (expressing future tense) and *jump* (expressing the semantic content JUMP). Without the lexical verb the verb phrase makes no sense: **kangaroo will* is not a grammatical sentence in English.

Auxiliaries typically express tense, aspect, mood, valency, voice or polarity of the verb phrase they belong to. Note that not all languages make use of auxiliaries; each of these grammatical categories may also be expressed morphologically or prosodically. Also, a language may mark part of a category with auxiliaries and part of it with some other method. Tense, for instance, is in English marked morphologically in the past (*walk* versus *walk**ed***) but with the help of an auxiliary in the future (*walk* versus ***will*** *walk*). For more on tense, mood, aspect, valency and voice, see Chapters 8 and 9. For more on polarity, see 12.1.1.2 and 12.1.2.1. For a very thorough survey and discussion of auxiliaries and auxiliary verb constructions, see Anderson (2009). For an article-length overview, see Anderson (2000). Some examples follow.

TENSE

(77) The man **will** write a song.

Tsou (Austronesian (Tsouic): Taiwan)

(78) **ta-ta** ɓoni ta tacɨmɨ
FUT-3SG.NOM AF.eat OBL banana
'He will eat a banana.' (Zeitoun et al. 1996: 39)

ASPECT

Kɔnɔ (Niger-Congo (Western Mande): Sierra Leone)

(79) à **á** tɛ̀ngbɛ̀ yón
3SG PFV basket spoil
'He has spoiled the basket.' (Kastenholz 2003: 33)

MOOD

(80) The man **must** write a song.

Peranakan Javanese (Austronesian (Javanese): Indonesia)

(81) Siti **harus** nomong Inggris
PN must speak English
'Siti must speak English.' (Cole et al. 2008:16)

VALENCY

Macushi (Cariban (Cariban): Brazil)

(82) arimaráákáˑ-yá pisaná ramá **máápiitíí-yá**
dog-AG cat see CAUSE-3SG.AG
'He caused the dog to see the cat' (Carson 1982:142)

VOICE

(83) The man **was** seen by the house.

Vietnamese (Austro-Asiatic (Viet-Muong): Vietnam)

(84) thuoc X **do** Y che nam 1973
medicine X PASS Y invent year 1973
'Medicine X was invented by Y in 1973.' (Keenan & Dryer 2007:341)

POLARITY

Finnish (Uralic (Finnic): Finland)

(85) **en** tule kotiin
NEG.1SG come home.ILL
'I won't come home.' (Sulkala & Karjalainen 1992:115)

In Examples (77) and (78) auxiliaries (*will* and *ta*) are used to mark tense (in both cases the future). In Example (79) the auxiliary *á* is used to mark perfective aspect. In Examples (80) and (81) the auxiliaries *must* and *harus* 'must' are used as modal markers. In Example (82) the auxiliary *máápiitíí* is used to mark the causative (*-yá* is an agreement marker). Incidentally, Example (82) also shows that an auxiliary may follow the verb it modifies. In Examples (83) and (84) the passive is marked by the auxiliaries *was* (an inflected form of *be*) and *do* 'PASSive' respectively. Example (84) also shows that an auxiliary need not be immediately adjacent to the verb it modifies. Finally, in Example (85) the auxiliary *en* marks negation.

6.2.2.6 *Conjunctions*

Conjunctions serve to connect entities (words, phrases or clauses). There are two types of conjunctions, those that assign the entities an equal status (**coordinating conjunctions**) and those that make one entity subordinate to another (**subordinating conjunctions**).

Examples of English coordinating conjunctions are *and*, *but* and *or*, as in, for example, *The boy **and** the girl played with each other*. English uses the same marker to connect verb phrases and clauses too, as in *The boy ate **and** drank* and *The boy sneezed **and** the girl coughed*. In other languages the choice of conjunction is dependent on what kinds of elements are to be conjoined. An example of a language that has three different conjunctions, one for connecting noun phrases (*iyo*), one for connecting verb phrases (*oo*) and one for connecting clauses (*-na* suffix), is Somali:

Somali (Afro-Asiatic (Eastern Cushitic): Somalia)

(86) a. rooti **iyo** khudrat
bread and fruit
'bread and fruit'

b. wuu cunay **oo** cabbay
FOC.3SG.M eat and drink
'He ate and drank.'

c. macallin-ku wuxuu joogaa dugsi-ga carruur-ta-**na** waxay
teacher-ART FOC.3SG.M be school-ART children-ART-and FOC.3PL
ku cayaarayaan dibed-da
PREV play outside-ART
'The teacher is in the school and the children are playing outside.'

(Haspelmath 2011c citing Berchem 1991: 324ff)

In Haspelmath's (2011c) sample of 301 languages, 161 (53.5%), English among them, use the same marker to coordinate noun phrases as they use to coordinate verb phrases, while 125 (41.5%), Somali among them, use different markers. The remaining 15 (5%) simply juxtapose the entities that are coordinated. For more on coordination, see 11.1.

Examples of English subordinating conjunctions are *because*, *if*, *that*, *while*, *although*, and so on, as in *She said **that** she would come* or *If it rains I won't go*. While most languages in Dryer's (2011j) sample have separate words[94] – a full 503 of 660 (76.2%) – some languages (64 or 9.7%) express subordination through suffixation. The subordinating conjunction may precede the clause it subordinates, as is the case in English.

(87) She said [**that** she would leave].

In (87) the subordinating conjunction *that* precedes the clause *she would leave*. This is the most common strategy in Dryer's sample, found in 399 (60.5%) languages. In 96 (14.5%) languages the subordinating conjunction follows the clause it subordinates. An example of such a language is Jamul Tiipay, where the subordinator *kenaach* 'because' follows the clause it subordinates:

94. Although not necessarily free forms, as clitics are also counted as 'words' Dryer's survey.

Jamul Tiipay (Hokan (Yuman): Mexico, USA)

(88) Juan may we-yiw xemaaw [xenu-ch **kenaach**]
PN NEG 3-come not be.sick-SS because
'Juan didn't come, because he is sick.' (Miller 2001: 256)

A very rare strategy, found in only 8 (1.2%) languages in Dryer's sample, is to have the subordinator appear inside the clause it subordinates, as is the case in Nkore-Kiga:

Nkore-Kiga (Niger-Congo (Bantoid): Uganda)

(89) [wa-ruhitsi **ku** a-za ku-taasya] a-shanga oburo ni-bu-sya
mr-hyena when he-go to-return he-find millet PC-it-cook
'When Brer Hyena starts to bring home (the cattle), he finds the millet cooking.'
(Taylor 1985: 26)

In Example (89) the subordinating conjunction *ku* 'when' appears inside the clause (*waruhitse aza kutaasya*) it subordinates. Some languages (93 or 14.1%) have mixed systems, in that they may allow for both preceding and following conjunctions or in that they have both separate words and suffixation to express subordination. For more on subordination, see 11.2.

CLAUSE INTRODUCERS

Urarina, a language isolate spoken in Peru, has, in addition to verbal inflection, a small class of function words that indicate the type of clause coming up. The introducer *ʤatera* ('how about…; what if…; etc.') indicates that a suggestive clause is coming up:

ʤatera hitarïi kaɯ tete-riɯ
how.about all here make-SUG.1SG
'What if I put all (those things) here?'
(Talking about items someone else had promised.) (Olawsky 2006: 267)

6.2.2.7 *Interjections*

Interjections are their own utterances that typically "express a speaker's current mental state or reaction toward an element in the linguistic or extralinguistic context" (Ameka 2006: 743). This is the only closed word class that is probably universal to all spoken languages. Interjections are typically used to express emotions (which includes the use of swear words), but there are other functions as well. Examples of some English interjections are *ouch!*, *yuck!*, *psst!*, and so on. Interjections are often, but by no means always, monomorphemic, do not take any morphological marking, and very often display sound sequences otherwise not typical for the language. For instance, in English words usually have to include at least one vowel. However, interjections such as *psst!* violate that pattern. A crucial feature of interjections is that they form independent

nonelliptical utterances, i.e. they are full utterances that do not constitute a shortened version of a longer utterance.

Interjections may have different functions. **Expressive interjections** indicate the speaker's mental state. These can be either emotive or cognitive. Examples of emotive interjections are the English *yuck!* 'I am disgusted' or *ouch!* 'I feel (sudden) pain', or the Swahili (Niger-Congo (Bantoid): Tanzania) *salala!* 'I am surprised' (Eastman 1992: 276). **Cognitive interjections** indicate the state of knowledge or thoughts of the speaker, as in the English *aha!* 'I understand' or the Italian *bo'!* 'I don't know'. **Conative interjections** are utterances directed at an addressee. Examples are the English *shh!* 'Be silent!' (which may also be called a volitive interjection) and the Italian *to'!* 'Take this!' (which may also be called a presentational interjection). **Phatic interjections** are used as communicative cues, for instance the English *uh-huh* 'I am following what you're saying' or such things as greetings (e.g. English *hello!* or Swedish *hej!* 'hello') and leave-taking (e.g. German *tschüss!* 'bye').

6.3 Parts-of-speech in sign languages

While cross-linguistic overviews on parts-of-speech systems in sign languages are very much in their infancy, it is safe to assume that sign languages, just like spoken languages first of all distinguish between open and closed word classes.[95] Secondly, it also seems safe to assume that the categories Noun and Verb (or **entity** and **event** signs in the terminology of Schwager & Zeshan 2008) are universal to sign languages, even though there might be significant overlap between them, much like we find in, for example, Nuuchahnulth. Thus in IPSL the sign for WORK can function either as a noun or a verb, as Example (90) shows:

IPSL (Sign Language: India, Pakistan)

(90) INDEX_1 WORK A-LOT
  a. 'I work a lot.'
  b. 'My work is a lot.'
  c. 'I have a lot of things to do (works).' (Schwager & Zeshan 2008: 513)

In Example (90) the sign WORK can refer either to the verb (event) 'to work' (a) or to the noun (entity) 'work' (b and c). However, as with spoken languages, there tends to be subtle differences between entity and event words, for example with respect to their morphology or syntax, which indicates that it makes sense to assume two different

95. As was mentioned in Section 5.3, sign languages make use of both derivation and compounding to form new words.

word classes, albeit with considerable overlap. In LSQ, for instance, word-mouthing is primarily associated with entity signs while expressive mouth movement is primarily associated with event signs (Voghel 2005). In ASL an event sign cannot pre-modify other signs and cannot combine with a quantifier (Schwager & Zeshan 2008).

Sign languages may also have modifiers (**property** signs in Schwager & Zeshan 2008), which may function either as adjectives, i.e. signs that modify entity signs, or as adverbs, i.e. signs that modify event signs. As in spoken languages, sign languages may differ in the scope and usage of these signs. Thus DGS property signs can be used in predicate functions, as modifiers of entity signs and as modifiers of event signs, while property signs in KK are only used in predicate functions.[96]

DGS (Sign Language: Germany)

(91) a. BIRD SMALL DEAD
'A small bird is dead.'

b. $POSS_1$ WIFE WORK GOOD
'My wife works well.' (Schwager & Zeshan 2008: 533)

KK (Sign Language: Indonesia)

c. $INDEX_1$ DEAF GATHER, GOOD
'It is nice when I get together with deaf people.'
(Schwager & Zeshan 2008: 533)

In Example (91a) the sign SMALL is used as an entity modifier (modifying BIRD) and the sign DEAD is used predicatively. In Example (91b) the sign GOOD is used to modify an event (WORK). Example (91c) shows the usage of the property sign GOOD in KK, which can only be used predicatively.

At this stage there is no cross-linguistic survey of the various closed classes in sign languages. IPSL, for instance, has a number of small, closed word classes, such as the non-manual signs expressing "yes", "no" and various adverbial meanings,[97] the classificatory stems consisting of handshapes that function as indicators of an entity engaged in an event (see 5.3 for an example in ASL), various functional particles, discourse markers, indexical signs (effectively pronouns and demonstratives) and auxiliaries. A cross-linguistic survey might find that classificatory stems, or classifiers, are very common to sign languages, possibly even universal (Sandler 2006).

---

96. Notice also that there is a certain overlap between entity and property signs in DGS: the sign GEHÖRLOS can mean both 'deaf' and 'deaf.person'. The sign DEAF in KK, however, always means 'deaf.person' (Schwager & Zeshan 2008: 525f).

97. Notice that this makes adverbs a closed class in IPSL, which also seems to be the case in ASL and Austrian Sign Language (ÖGS) (Schalber & Grose 2008).

Auxiliaries, which tend to carry grammatical information on agreement and valency (see, for example, Müller de Quadros 2008), might also turn out to be common across languages. Pronouns are typically expressed indexically, typically by pointing with the index finger (Schwager & Zeshan 2008). While interjections seem universal to spoken languages, there is at this stage no cross-linguistic study available on the equivalent to interjections in sign languages.

A survey of numeral signs cross-linguistically might show prevalence for manual counting, with the digits of the hand serving for the numerals one to nine.[98] However, it is likely that there will be differences in exact handshape, both between languages, as is the case between Catalan Sign Language (LSC) and Spanish Sign Language (LSE), both used in Spain, (Fernández-Viader & Fuentes 2008), and between varieties of the same language, as is the case in NZSL (McKee et al. 2008).

6.4 Summary

All languages, both spoken and signed, have two basic types of word classes (parts-of-speech), open classes and closed classes. The open classes, typically lexical classes, freely allow productive additions of new words to them, while closed classes, typically functional classes, do not readily permit additions. The maximum number of open classes a language can have is four: nouns, verbs, adjectives and adverbs. While there might be considerable overlap between the characteristics of nouns and verbs in some languages, it seems universal that there is never complete overlap. Languages differ as to whether adjectives and adverbs form separate, open classes in their system.

Languages differ radically as to how many and which closed word classes they might have, but typical categories include pronouns, adpositions, conjunctions, auxiliaries and articles. Interjections form the only closed word class that seems to be universal in spoken languages. There is a correlation between the amount of synthesis allowed in a language and the amount of closed class items it has.

Numerals may have formal affinities with either nouns or verbs, or constitute their own word class. Languages differ with respect to their numeral base, and may also have mixed systems, or may use body-parts to convey a numeral expression. There are also languages with no numerals at all.

Sign languages also organize their lexicon into open and closed word classes. While no major cross-linguistic survey for sign language parts-of-speech is available at this date, it seems that sign languages make use of rather similar types of classes. The open word classes for sign

98. But cf. NZSL, where one variant for '8' is expressed with a single closed fist held towards the middle of the upper chest (McKee et al. 2008: 302).

languages are entity signs (nouns), event signs (verbs) and property signs (modifiers functioning as either adjectives or adverbs or both). Closed classes may be, for example, pronouns (typically indexal), auxiliaries, classifiers and numerals. Classifiers seem to be universal to sign languages.

6.5 Keywords

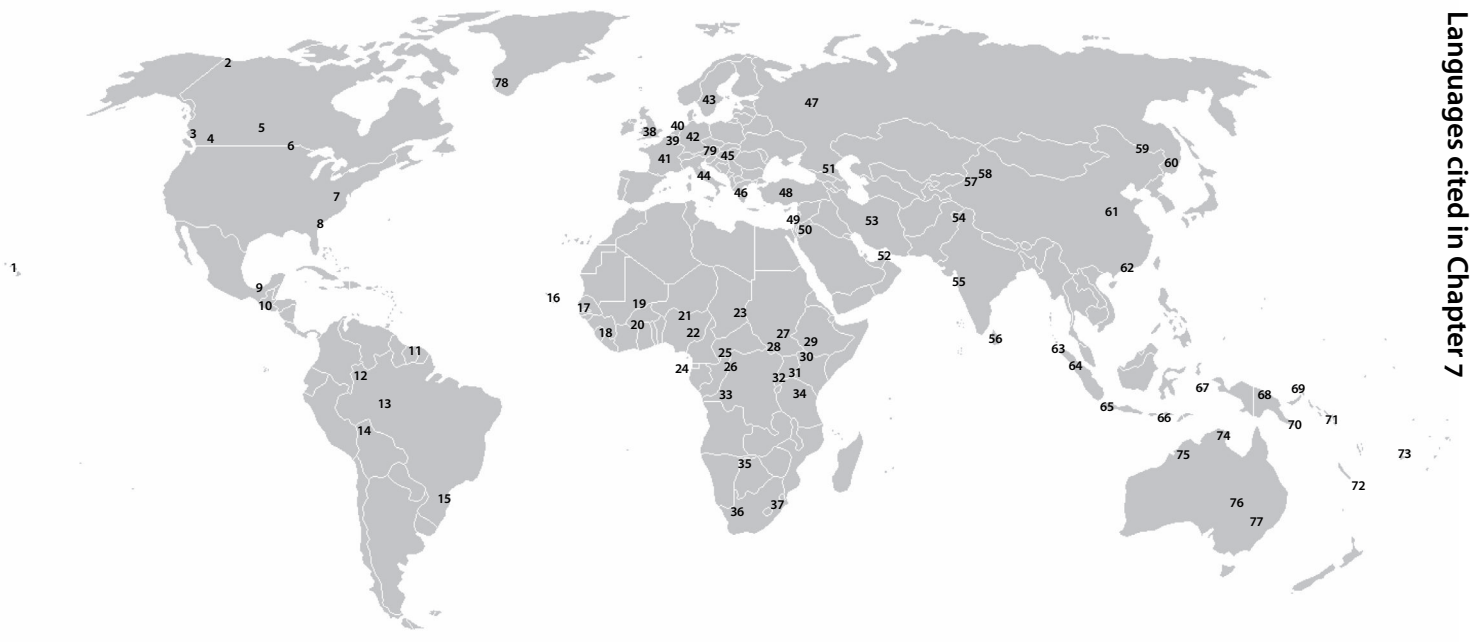
compounding
content word
derivation
function word
incorporation
lexical/functional classes
open/closed word classes
parts-of-speech
word formation

6.6 Exercises

1. What is **noun incorporation**?
2. Which lexical classes are languages most likely to have?
3. Define **inclusive** and **exclusive** pronouns. How do pidgin and creole languages pattern in comparison to non-creoles with respect to inclusive/exclusive differentiation?
4. How do sign language entity and event signs compare to spoken language parts-of-speech categories? State whether there is a clear-cut distinction between the two types of sign and how that compares to spoken languages.
5. Is the following statement true or false? Motivate your answer.

   Languages with predominantly analytic constructions make more use of function words than languages with predominantly synthetic constructions.

Languages cited in Chapter 7



1 Hawai'i Creole English
2 Eskimo Pidgin
3 Chinuk Wawa
4 Shuswap
5 Cree, Plains
6 Michif
7 American Sign Language
8 Gullah
9 Chontal Maya
10 Tzutujil
11 Ndyuka
12 Dâw
Hup
13 Pirahã
14 Cayuvava
15 Xokléng
16 Cape Verdean Creole of Brava
Cape Verdean Creole of Santiago
17 Banyun
18 Kisi
19 Jamsay
20 Dagaare
21 Hausa Sign Language
22 Angas
Fula, Nigerian
23 Maba
24 Fa d'Ambô
25 Sango
26 Lingala
27 Nuer
Shilluk
28 Dinka, Agar
29 Bayso
30 Maasai
31 Nandi
32 Nkore-Kiga
33 Kikongo-Kituba
34 Burunge
Iraqw
35 Ju|'hoan
36 Korana
37 Fanakalo
38 English
39 Dutch, Zeeuws
40 Ducth Sign Language
41 French
42 German
German Sign Language
43 Swedish
Swedish Sign Language
44 Italian
Italian Sign Language
45 Hungarian
46 Greek
47 Russian
48 Turkish
49 Israeli Sign Language
50 Jordanian Sign Language
51 Chechen
52 Khwarshi
Lak
Lezgian
Tabassaran
53 Persian
54 Kashmiri
55 Korlai
56 Sri Lanka Portuguese
57 Tocharian B
58 Tocharian A
59 Chinese Pidgin Russian
60 Udihe
61 Chinese, Classical
62 Chinese Pidgin English
63 Minangkabau
64 Kawi
65 Indonesian
66 Kambera
67 Larike
Ambon Malay
68 Manam
Mehek
Yimas-Arafundi Pidgin
69 Lihir
Nalik
Sursurunga
70 Gapapaiwa
71 Lavukaleve
72 Norf'k
73 Fijian
Pidgin Hindustani
74 Anindilyakwa
75 Ungarinjin
Worora
76 Australian Sign Language
77 Ngiyambaa
78 Greenlandic, West
79 Austrian Sign Language

Chapter 7

Nominal categories and syntax

Nouns may be subcategorized according to different criteria, for instance if they can be counted or not. When several elements form a unit with a noun we have a noun phrase. Noun phrases may be specified for whether they refer to one or several entities, whether they belong to any specific class or which relation they have to the other constituents in the phrase. The elements in the phrases tend to have an internal organization, with a head and one or more dependent(s). In this chapter I very briefly show some of the main subcategorizations for nouns before giving a brief overview of some major noun phrase categories (7.1). In 7.1.1 I present various number values found in the languages of the world, while in 7.1.2 I present some of the gender or noun class systems found cross-linguistically. In 7.1.3 I discuss case and give a necessarily brief overview of some of the most common cases. Section 7.2 deals primarily with the inner organization of noun phrases: I first define the terms head and dependent and show how languages vary with respect to locus of marking (7.2.1) and then discuss the difference between government and agreement (7.2.2). Section 7.3 gives an overview of noun phrase operations in sign languages.

7.1 NP categories

The **noun phrase** (**NP**) is the entity which functions as an argument, such as a subject or an object, in a sentence. It consists minimally of a noun or a substitute for a noun (for example a pronoun), but may also consist of several words that belong together as a phrase, where the noun or its substitute constitutes the core element (the **head**). For instance, the sentence *Kangaroos jump* consists of the noun phrase *kangaroos* and the verb phrase *jump*. The noun phrase here consists of only one word, the noun *kangaroo*. However, the sentence *The kangaroo is jumping* also only consists of one noun phrase (*the kangaroo*) and one verb phrase (*is jumping*), even though each of these phrases contain two words. Examples of various noun phrases, here underlined, are *He jumped*, *The two of them were jumping*, *The happy kangaroo on the other side of the fence was jumping*, *The man who won the lottery danced*.

Nouns can be divided into a number of grammatical subclasses, usually based on common semantic properties of the noun. While the following subclasses do not constitute an exhaustive list, they are commonly found in the languages of the world.

WHAT'S IN A NAME?

In Mehek (Sepik (Tama Sepik): Papua New Guinea) each person's name has four counterpart forms: one of the forms is used when the person is being reprimanded, the other of the forms is used when the person has done something good or has returned from a journey, the third form is a whistle call used to call the person if s/he is out of visual range, and the fourth form is a song used as a form of greeting or praise. (Adam Hatfield, p.c.)

Proper names, or **proper nouns**, in contrast to **common nouns** (all others), are typically used to refer to specific individuals or places. Examples of proper nouns are *Peter*, *Mr. Smith*, *London*, *Italy*, *Mississippi River*, *Mount Kilimanjaro*. While common nouns can be specified for definiteness, proper nouns usually cannot. Thus we might need to specify *the chair* (definite) as opposed to *a chair* (indefinite), but we rarely need to specify **the Italy* as opposed to **an Italy*, since the inherent meaning of the proper noun is that it is a specific (definite) individual or place we are referring to. There are, however, occasions when the use of definite articles with proper nouns might be justified, for example *I did talk to a Jenny at the conference, but she didn't look like the Jenny you described*, where the implication is that we need to distinguish between several individuals with the same name.

Common nouns are often divided into **count nouns**, i.e. such nouns that can be counted, and **mass nouns** (or **noncount nouns**), i.e. such nouns that cannot be counted. Examples of count nouns are *tree*, *star*, *property*. Examples of mass nouns are *air*, *sand*, *anger*. Count nouns may refer to sets containing one or more separable entities of the same: *a tree* (/*star*/*property*), *two trees* (/*stars*/*properties*), *many trees* (/*stars*/*properties*), while mass nouns generally do not (compare **a sand*, **two sands*, **many sands*) but tend to refer to continuous entities that are not easily separable. Greater or smaller quantities of mass nouns may then be expressed by other quantifiers, as in English *some* and *much*: *some air* (/*sand*/*anger*; compare ??*some tree*/*star*/*property*), *much air* (/*sand*/*anger*; compare ??*much tree*/*star*/*property*).

Languages may also subclassify nouns according to whether they can or cannot be possessed, whether nouns must obligatorily be possessed as opposed to possession being only a non-obligatory option, or whether possession can be transferred or not (so-called **alienable** versus **inalienable possession**). An example of a language with a subcategory of nouns that cannot be possessed is Tzutujil (Mayan (Mayan): Guatemala) (Dayley 1981: 199ff) where nouns referring to natural phenomena, wild animals or people cannot be possessed, for example *juyuʔ* 'mountain', *b'ajlam* 'jaguar', *q'isaaneel* 'witch' – something like **nuujuyuʔ* 'my mountain', **nuub'ajlam* 'my jaguar' or **nuuq'isaaneel* 'my witch' is not possible while, for example *nuutz'iiʔ* 'my dog', *nuukaab'* 'my raw sugar, honey' and *nuutiiʔ* 'my meat' are quite grammatical and accepted forms. In Tzutujil there is also a subcategory of nouns that must be possessed,

inherently possessed nouns (also called **bound nouns** or, which is more precise, **obligatorily possessed nouns**, cf. Nichols & Bickel 2011b), where the item referred to has to be specified with respect to whom it belongs. In other words, expressing the noun without a possessor marker specifying who the item belongs to would be ungrammatical. These are usually nouns that have some inherent relationship with something else, such as kinship terms, body parts, tools, abstract nouns, and so on, for instance *nuumaam* 'my grandchild' (but **maam* 'grandchild') (Dayley 1981: 205). This is often called inalienable possession. However, it makes sense to differentiate between inherent possession, where the owner must be specified, and inalienable possession, where the owner of an item does not necessarily have to be specified for the utterance to be grammatical. The difference between alienable and inalienable possession lies in *how* the possession is expressed. Possession of alienable items, that is, such items where ownership can be transferred (for example worldly goods) is expressed differently from possession of inalienable items, i.e. such items that cannot be transferred (for example body parts or kinship terms). Thus in Nalik, inalienable possession has to be marked as in Example (92a), while alienable possession has to be marked as in Example (92b):

Nalik (Austronesian (Oceanic): Papua New Guinea)

(92) a. a langa-go ka burus
ART ear-my 3 hurt
'My ear hurts.'

b. ka zaxot a buk surugu
3 like ART book of.I
'He wants my book.' (Volker 1994: 178f)

In (92a) possession is expressed with the suffix *-go*, while in (92b) possession is expressed with the free word *surugu*. Languages that distinguish between alienable and inalienable possession thus have at least two ways of expressing possession. Languages with inherently possessed nouns, on the other hand, can make do with only one way of expressing possession – the crucial thing about inherently possessed nouns being simply that the owner has to be specified, which may well be the same kind of possessive expression as expressing ownership of nouns that are only optionally possessed. This is shown in the examples from Tzutujil mentioned above, where possession may be expressed with the suffix *nuu-* irrespective of whether the noun is inherently possessed or only optionally possessed.

Languages may subcategorize noun phrases (i.e. both nouns and their substitutes) according to **animacy**. The most basic animacy distinction would be between animate and inanimate nouns, such as between *dog* and *stone*. English reflects this distinction in that we normally would use *he* (or *she*) when referring to a dog but *it* when

referring to a stone. However, languages may make further distinctions. For instance, in English proper names and kin terms can usually not be combined with the indefinite article: compare *Peter* but **a Peter*, *a friend of mine* but **a mother of mine*. We might say here that English displays an **Animacy Hierarchy** along the lines of proper nouns and kin terms > other animates, where '>' indicates that proper names and kin terms are higher up the hierarchy in terms of definiteness (they are inherently definite) than other animates.[99] English also differentiates between humans versus non-human animates and inanimates with respect to which relative pronoun is used: compare *the book which I read*; *the bird which flew away* (not **the book whom I read* or **the bird who flew away*) and *the girl whom I saw* (not **the girl which I saw*). We could, in fact, refine the English animacy hierarchy to something like proper nouns and kin terms > humans > other animates >inanimates. Languages might code things according to even more precise distinctions, also taking into account whether the NP refers to the speaker, the addressee, or some third person, apart from the other distinctions mentioned. A more detailed Animacy Hierarchy would then be as in Figure 7.1:

1st person > 2nd person > 3rd person > proper noun/kin > human > animate[100] > inanimate
speaker addressee

Figure 7.1 The Animacy Hierarchy.

In Figure 7.1 the further left the slot is, the more 'animate' the NP is. The term 'animate' is slightly misleading here, since obviously the noun *mother* does not indicate a living being more animate than the noun *woman* or the noun *dog*. However, if we think of the hierarchy as structured according to sociocentric orientation (Whaley 1997: 172) or as arranged according to a principle of empathy (Payne 1997: 151) the hierarchy becomes more transparent. In essence it simply captures the notion that we humans tend to identify most with ourselves (I am most aware of myself), then with the addressee (I am very aware of you, whom I am talking to), then with some other third person (I am quite aware of him/her/them, whom I am talking about), then, in continued descending order, with John and my sister (proper names/kin), with the boy in the yard (human NP), with the cow in the field (animate NP), and, finally, the entity that we tend to identify the least with would be the stone (inanimate NP). In fact we

99. This could also be termed Topicality Hierarchy or a hierarchy of topic-worthiness, since the hierarchy in principle depicts the order in "which noun phrases are more likely to occur as topics" in discourse (Comrie 1989: 198).

100. It would be more accurate to have the term 'non-human animate' here, as it denotes all animals (or animate beings) that are not human.

shall see throughout the remainder of this book that languages often have to code things differently in their grammar depending on where the noun phrase is placed in the Animacy Hierarchy.

7.1.1 Number

7.1.1.1 *Number values*

Most languages have a grammatical way of expressing whether one or more real world entities are referred to.[101] In English, for example, there are usually two forms to choose between, one a base form indicating **singular** (one entity) and the other the base form modified in some way indicating **plural** (more than one entity), as in *chair* versus *chair*s. There are languages, however, where the base form of the noun actually does not give any information as to the number of entities involved, called **general number** in Corbett (2000: 9ff). While this is cross-linguistically rare, it can be found in Bayso, where *lúban* 'lion' is not specified for number and thus means that it could be one or more lions, as opposed to *lúbantiti* which means exactly one lion, and *lúbanjool*, which means many lions, as in Example (93):

Bayso (Afro-Asiatic (Eastern Cushitic): Ethiopia)

(93) a. lúban foofe
lion.GENERAL watched.1SG
'I watched lion' (number not specified; could be one or more than one)

b. lúban-titi foofe
lion-SG watched.1SG
'I watched a lion' (exactly one)

c. lúban-jool foofe
lion-PL watched.1SG
'I watched lions' (many)[102] (Corbett 2000: 11 citing Dick Hayward, p.c.)

The most common system is to have a two-way contrast between singular and plural, as in English. However, it is not an absolute universal to express plurality: there is at least one known language, Pirahã, which actually does not have any grammatical number (Everett 2005).[103] A sentence like in Example (94) simply does not specify the NPs for number, as can be seen from the various translations it can get:

---

101. For a discussion on verbal number, expressing multiple events, see Corbett (2000: 243ff) and 8.4.

102. Since Bayso also has paucal (for example *lúbanjaa* '(a few) lions'), the plural refers to more than just a few of an entity. See below.

103. It seems that Kawi or Old Javanese (Austronesian (Javanese): Indonesia) and Classical Chinese (Sino-Tibetan (Chinese): China), both extinct, also lacked grammatical number; see Corbett (2000: 51) for references.

Pirahã (Mura (Mura): Brazil)

(94) hiatíihi hi kaoáibogi bai-aagá
Pirahã 3 evil.spirit fear-COP
'The Pirahã people are afraid of evil spirits.' or
'The Pirahã people are afraid of an evil spirit.' or
'A Pirahã is afraid of evil spirits.' or
'A Pirahã is afraid of an evil spirit.' (Everett 2005: 623)

In Example (94) neither *hiatíihi* 'Pirahã person / Pirahã people', nor *hi* '3rd person (singular or plural)', nor *kaoáibogi* 'evil spirit(s)' are specified for number.

It is not uncommon for languages to have yet another number value, the **dual**, which specifies that two (and exactly two) entities are referred to. An example of a language with the three number distinctions singular, dual and plural is Lavukaleve:

Lavukaleve (Solomons East Papuan (Lavukaleve): Solomons Islands)

(95)

SINGULAR	DUAL	PLURAL	
filifil	filifili**l**	filifili**mal**	'nail'
mulukita	mulukita**ul**	mulukita**vil**	'orange'

(Terrill 1999: 96f)

In Lavukaleve nouns are inflected for three numbers, with the dual and the plural overtly marked through suffixes (bolded), as shown in (95). Notice that in languages with these three number distinctions the plural necessarily means 'more than two', since the dual specifies for '(exactly) two of X'. Examples of pronouns with the number distinctions singular, dual and plural were given in Section 6.2.2.1.

TWO OF THE SAME AND NATURAL PAIRS

Tocharian (Indo-European (Tocharian)) is a cover term for two distinct languages, Tocharian A (East Tocharian or Agnean) and Tocharian B (West Tocharian or Kuchean). The documents we have for these languages, mostly translations of Buddhist texts, are from the 6th to 8th centuries. Tocharian A texts have only been found around Turfan and Quārāšahr (the ancient Agni – hence the alternative name Agnean) while Tocharian B texts have been found from Tumšuq in the west to Turfan in the east, especially in Kuča (hence the alternative name Kuchean), along the eastern branch of the Silk Road in present day Xinjian (western China).

While both Tocharian languages retained the Indo-European dual, they also developed another kind of dual, the **paral** (also called **ambal**), used only for naturally occurring pairs, as in, for example, 'both eyes':

Tocharian A: *aśäṃ* Tocharian B: *eśane*

According to *Das grammatische Raritätenkabinett* in Konstantz (http://typo.uni-konstanz.de/rara/intro/), this is not known for any other language (living or extinct) in the world. For more information on Tocharian, see Krause & Slocum (2007–2010).

Tough it is rather uncommon, languages may have a further number distinction, the **trial**, specifying for three (and exactly three) entities, giving a four-way distinction of singular, dual, trial and plural. There are no known languages that have trial but not dual (yielding a three-way system with singular, trial, plural). An example of a language with a trial number category is Larike.

Larike (Austronesian (Central Malayo-Polynesian): Indonesia)

	SINGULAR		DUAL		TRIAL		PLURAL	
1-INCL	–		itua	we two	itidu	we three	ite	we many
1-EXCL	aʔu	I	arua	we two	aridu	we three	ami	we many
2	ane	you	irua	you two	iridu	you three	imi	you many
3	mane	he/she	matua	they two	matidu	they three	mati	they many

(Laidig & Laidig 1990: 90)

The Larike free pronouns, shown above, tend to only be used for human referents and so only one form is given for the 3rd person (with pronominal affixes a formal distinction is made between humans and nonhumans). In a sentence like (96) the 3rd person can thus only refer to persons.

Larike (Austronesian (Central Malayo-Polynesian): Indonesia)

(96) matidu-tue au-huse nusa
3TRI:S-live at-there island
'Those three live on the island over there.' (Laidig & Laidig 1990: 96)

Notice that the trial, when it occurs, is employed in the pronominal system only; there are, as yet, no known languages that employ the trial for nouns. In some languages the form termed 'trial' may actually refer to more than three, and would, as such, be better termed **paucal** ('a few'). An example of a language with a number category that means roughly "three or a small group" is Manam (Austronesian (Oceanic): Papua New Guinea) where the pronominal forms termed 'triple' is "three or a few in a group" (Turner 1986: 66). A more precise way of describing the numbers of Manam would therefore be to say that it has a four-way system of singular (one X), dual (two X), paucal (a few X but more than two) and plural (more than a few X). Most languages with a paucal number category also have a dual category, but Bayso, mentioned above, has a three-way distinction of singular (one X), paucal (a few X) and plural (more than a few X). The most complex systems known have five different number values. Lihir, for example, has singular (one X), dual (two X), trial (three X), paucal (a few X but more than three) and plural (more than a few X):[104]

104. Notice, however, that the exact status of the trial is not known; it could be that it refers to exactly three, in which case it is a genuine trial, but it could also mean that it refers to 'three or a few', like the 'triple' in Manam, in which case it might be a type of paucal.

Lihir (Austronesian (Oceanic): Papua New Guinea)

	SINGULAR	DUAL	TRIAL	PAUCAL	PLURAL
1-INCL	–	kito	kitol	kitrahet	giet
1-EXCL	yo	gel	getol	gehet	ge
2	wa	gol	gotol	gohet	go
3	e	dul	dietol	diehet	die

(Corbett 2000: 25 citing Malcolm Ross, unpublished fieldnotes)

There have been claims of languages with a number value 'quadral' (four of X), such as Sursurunga (Austronesian (Oceanic): Papua New Guinea) (Hutchisson 1986), which also has a five-way number distinction. However,

> [s]ince plural pronouns are never used with relationship terms [i.e. terms reflecting kinship], the use of these terms skews number reference for both trial and quadral forms (although not for dual), so that trial comes to mean a minimum of 3, and quadral a minimum of 4. (Hutchisson 1986: 10)

Hence Sursurunga number values labelled 'trial' and 'quadral' do not refer to exactly three and four units respectively. A more exact way of labelling the Sursurunga number system would therefore be with the distinctions singular (one X), dual (two X), paucal (three or so X/a few X), greater paucal (four or so X/slightly more than a few X) and plural (many X).

To distinguish between a 'normal amount' and a 'greater than normal amount', as is done in the Sursurunga paucal is very rare and, if anything, is found with the plural yielding a **greater plural** (or **global plural**). The distinction would then be something like 'many' and 'very many indeed'. An example of a language with greater plural is Banyun (Niger-Congo (Northern Atlantic): Senegal); compare *bu-sumɔl* 'SG-snake' with *i-sumɔl* 'PL-snakes' and *ba-sumɔl* 'GR.PL-snake', where the difference between *i-sumɔl* and *ba-sumɔl* is that the latter "is used when the number cannot be counted or the speaker feels it unnecessary" (Corbett 2000: 31).

7.1.1.2 *Obligatoriness*

Languages differ as to whether it is actually necessary for the speakers to mark number and if so, which nominals have to be marked. In English all sorts of nominals have to be marked for number, whether they denote humans, animals or inanimate objects. That is, grammatical marking of plural is generally obligatory and is done on all kinds of nominals. This is the most common pattern, comprising 133 of 291 (or 45.7%) in Haspelmath's (2011d) sample, which was coded for whether plural marking was obligatory, optional, or lacking altogether, as well as whether the nominals involved were human animates or discrete inanimates (i.e. count noun inanimates). In 55 languages (18.9%) plural marking is optional for all nominals, while in 15 languages

(5.2%) plural marking is obligatory for human and only optional for inanimate nominals. A number of languages in the sample only mark plurality on human nouns: in 40 (13.7%) of them plural marking is obligatory while in 20 (6.9%) it is optional. 28 (9.6%) languages of the sample lack grammatical marking of plural. There are no languages in the sample that mark plurals (optionally or obligatorily) only on inanimate nouns but not on human nouns, nor are there any languages where plural marking is obligatory with inanimates but only optional with humans. In other words, Haspelmath's (2011d) sample exhibits a pattern that is in accordance with the Animacy Hierarchy: if a language has a split in obligatoriness of plural marking, where some nouns have to be marked for plural and others do not, it is safe to assume that it is the human nouns that will have obligatory marking and the inanimate nouns that will have optional marking.

The languages in the APiCS differ from the languages in Haspelmath's sample in their pattern with respect to obligatoriness in plurality marking. While the proportion of languages lacking a nominal plural is similar to that in Haspelmath's sample, the proportion of languages for the other values differ, as summarized in Table 7.1:

Table 7.1 Comparison between WALS and APiCS languages for occurrence of nominal plural marking. Adapted from Haspelmath (2011d) and Michaelis et al. (2013: feature 22). Absolute numbers in parentheses.

	Value	WALS		APiCS[105]	
1.	No nominal plural	9.6%	(28)	6.6%	(5)
2.	Only human nouns, optional	6.9%	(20)	5.3%	(4)
3.	Only human nouns, obligatory	13.7%	(40)		(0)
4.	All nouns, optional	18.9%	(55)	67.1%	(51)
5.	All nouns, optional in inanimates	5.2%	(15)		(0)
6.	All nouns, obligatory	45.7%	(133)	21.1%	(16)
	Total		291		76

Table 7.1 shows that it is more likely that a pidgin or creole language allows plural marking on all nouns, but that it is optional, while it is more likely for non-creoles to have obligatory plural marking on all nouns.

105. The languages for value 1 are Chinuk Wawa, Eskimo Pidgin (Pidgin (Eskimo-lexified): Canada), Fanakalo, Korlai and Pidgin Hindustani; the languages for value 2 are Chinese Pidgin English (Pidgin (English-lexified): China), Chinese Pidgin Russian (Pidgin (Russian-lexified): China), Fa d'Ambô (Creole (Portuguese-lexified): Equatorial Guinea) and Yimas-Arafundi Pidgin.

7.1.1.3 *Associative plural*

Associative plural (also called **group plural**, among other terms)[106] basically means 'X and those associated with X'. It is typically used with nouns referring to humans, usually proper names, kinship terms, titles and occupations. Strictly speaking this is not a number value or a category of number, although it is almost always used in combination with number. For a discussion on the difference between associatives and number, see Corbett (2000: 101ff, especially 110f).

An example of a language with associative plural is Hawai'i Creole English with the clitic *-dem*, as in *John-dem* 'John and them/John and his friends/John and those associated with him' or *ma faðɛ-dem* 'my father and them/my father and those associated with him'. This differs semantically from **additive plurals**, i.e. those kinds of plurals that simply mean more of the same, which is what the discussion above has focussed on. A construction like *John-dem* does not mean *'many Johns', nor does *ma faðɛ-dem* mean *'my many fathers'. Instead the associative denotes a group of unspecified individuals that are in some way associated with a named referent. I will follow Daniel & Moravcsik (2011) and term the named referent (e.g. *John* or *my father*) the **focal referent**, and term the group of individuals that are associated with the focal referent **associates**. Notice that the associates in associative plural may form a group of different individuals: those associated with the focal referent may be of different genders or ages, or have different kinds of kinship relations to each other. The associates in an associative plural may thus form a heterogeneous group, as opposed to the nouns referred to in an additive plural (*girls* means 'many young female humans', *mothers* means 'many females who have children'). For a very thorough discussion on the semantic properties of associative plurals, see Moravcsik (2003).

Associative plurals are most commonly formed with proper names and kinship terms. In their sample of 237 languages, Daniel & Moravcsik (2011) found that the vast majority, 200 languages (or 84.4%), have associative plural. English, like most other Western European languages, belongs to the minority group that does not have any associative plural. The most common strategy, found in 105 languages (or 44.3%) of the sample, is for a language to have the same form for associative plural as for additive plural. An example of such a language is Udihe; compare Examples (97a and b):

Udihe (Altaic (Tungusic): Russia)

(97) a. tege — tege-ziga
gown — gown-PL
'gowns'

106. For this discussion I collapse everything that is 'more than one' into the term plural. In other words, I am here not differentiating between dual, trial, paucal and plural.

b. Guatu Guatu-ziga
PN PN-PL
'Guatu and those associated with him.'
(Nikolaeva & Tolskaya 2001: 115f)

In Example (97) the same form is used to mean 'more of the same' (*tegeziga* 'gowns') as well as 'those associated with X' (*Guatuziga* 'Guatu and those associated with him').

It is also quite common to have different forms for additive and associative plural, a strategy found in 95 (or 40.1%) of the languages in Daniel & Moravcsik's sample. This is, for instance, the case in Hawai'i Creole English. The above mentioned clitic *-dem* can only mean 'and those associated with X'. A construction like **buk-dem* ('the book and those associated with it') is not possible, unless used in a fairy-tale sense to give the book its own cartoon character. Crucially, **buk-dem* can never have the additive plural meaning of 'books'.

The figures for the languages in APiCS are at the time of writing still temporary for this feature, but they seem to indicate a different pattern from the languages in WALS (Michaelis et al. 2013: feature 24). In the APiCS sample the picture seems less radical in that only slightly over half of the languages have associative plural (compared to the 84.4% in WALS) while some two fifths lack the associative plural (compared to the 16.4% in WALS). It thus seems rather more common for a non-creole to have an associative plural marker than it is for a pidgin or creole language.

7.1.2 Noun classes (gender) and classifiers

7.1.2.1 *Gender*

Gender or **noun class** is a grammatical classification for nouns or substitutes for nouns (such as pronouns). For very thorough discussions of gender, see Corbett (1991 and 2007), which this section relies heavily on. Those of us mainly used to Indo-European languages might think of 'gender' as a classification largely corresponding to sex, with the distinctions 'female', 'male' and possibly a non-sex category 'neuter', but as we shall see this is by no means a universal system. The actual evidence for a gender or noun class system lies outside the noun itself, as it is only reflected by way of agreement in the words associated with the given noun. In other words, it is not possible to determine which gender or noun class a word in a language with such a system belongs to by just looking at it. Rather, the gender or noun class has to be established through the different agreement patterns that different genders or noun classes take. In German, for instance, we have three genders or noun classes, masculine, feminine and neuter. However, the gender of the words *Stuhl* 'chair', *Blume* 'flower' and *Buch* 'book' can only be determined by some kind of agreement form outside the noun, such as the definite article (***der*** *Stuhl* 'the.M chair', ***die*** *Blume* 'the.F flower', ***das*** *Buch* 'the.N book') or the form of the adjective (*ein rot-**er** Stuhl* 'a red-M

chair', *eine rot-**e** Blume* 'a red-F flower', *ein rot-**es** Buch* 'a red-N book').[107] Agreement targets may be adjectives, various kinds of pronouns, articles, possessives, numerals, verbs, participles, adverbs, adpositions and in Zeeuws Dutch even complementizers, as shown in Example (98).

Dutch, Zeeuws (Indo-European (Germanic): The Netherlands)

(98) a. …dank (ik) kommen
that I come
b. …daj (gie) komt
that you.SG comes
c. …datje (jij) komt
that he comes
d. …dase (zie) komt
that she comes
e. …dat (et) komt
that it comes
f. …dame (wunder) kommen
that we come
g. …daj (gunder) komt
that you.PL come
h. …danze (zunder) kommen
that they come (Bennis & Haegeman 1984: 41)

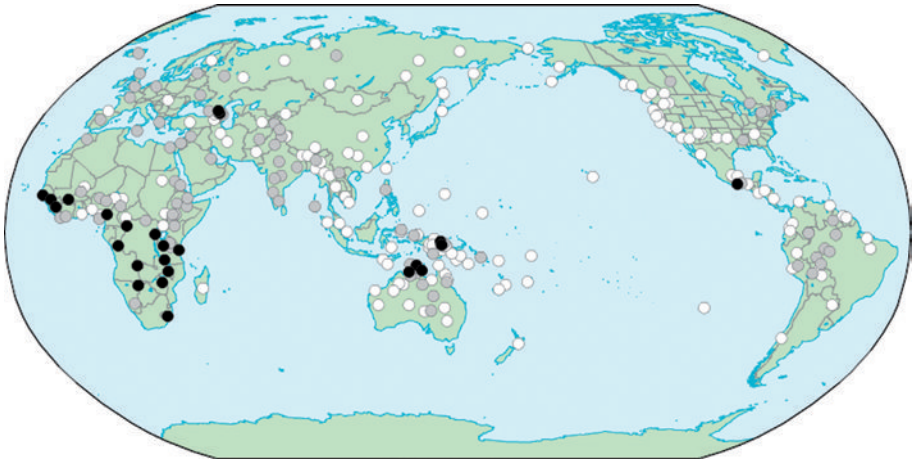
In Example (98) the complementizer takes different forms depending on the person, number, and in the case of the third person singular, the gender of the subject pronoun in the clause: *dank* 'that.I'; *daj* 'that.you' (either singular or plural); *datje* 'that.he', *dase* 'that.she' or *dat* 'that.it'; *dame* 'that.we' and *danze* 'that.they'. The pronoun is usually omitted, but may appear if the speaker wishes to stress it, which is why it is given in parentheses. For more on agreement see 7.2.2.

It is common to include so-called **pronominal gender systems** as an instance of a gender agreement system. These are instances when the gender of the free pronoun is determined by the word it refers to (its **antecedent**), as in English *The man… he…* or *The woman… she…* where the pronoun has to be in the masculine (*he*) when the antecedent is a noun referring to something male (*the man*) and in the feminine (*she*) when the antecedent is a noun referring to something female (*the woman*). English is an example of a language with a pronominal gender system, where the free pronouns constitute the only evidence for gender. This is cross-linguistically rare (Corbett 2011a).

107. In actuality the indefinite article *ein*/*eine* also shows agreement, but is ambiguous for masculine and neuter without an adjective.

It might seem like a given that gender systems will somehow be sex-based in that males belong to the masculine gender, females to the feminine, and in case of a third gender, that non-sex items will belong to the neuter. However, while sex-based gender systems are much more common than non-sex based systems, they are not universal: of the languages in Corbett's (2011b) sample with a gender system, 75% were sex-based while 25% were not. The latter group of languages all base their gender system on some form of animacy (Corbett 2011b). An example of a language with a non-sex-based gender system is Plains Cree (Algic (Algonquian): Canada), where there are two genders, one for animate nouns and one for inanimates (Wolfart 1973: 20). This means that both male and female humans and animals belong to the same gender, the animate one. Kisi also has a non-sex-based gender system with its four genders denoting animates, inanimates, collective plants and grains, and liquids, respectively (Childs 1995: 148). To those of us who are most used to sex-based systems, like those found in French or German, this might seem counterintuitive, as might the concept of more than three genders. But etymologically 'gender' derives from Latin *genus* 'kind, sort'; gender and noun class thus mean the same thing, namely the categorization of nouns into different classes.

Gender or noun class systems may have a minimum of two genders, as is the case in French, or as many as around 20 genders, as is the case with some dialects of Nigerian Fula (Niger-Congo (Northern Atlantic): Nigeria). However, more common cross-linguistically is to not have any gender system at all, as can be seen on Map 7.1. In Corbett's (2011a) sample of 257 languages, more than half (145 languages or 56.4%) lack gender.



Map 7.1 Number of genders or noun classes. White dots indicate languages with no gender or noun class; grey dots indicate languages with 2–4 genders; black dots indicate languages with 5 or more genders. Modified from Corbett (2011a). For a full legend, see http://wals.info/feature/30A.

Of those languages that do have gender, the most common is to have only two, found in 50 (19.5%) languages. An example of a language with two genders is, as mentioned, French, with feminine (e.g. ***la*** *table* 'the.F table') versus masculine (e.g. ***le*** *livre* 'the.M book'). Swedish is another example of a language with two genders, common (***en*** *bok* 'a.C book') and neuter (***ett*** *träd* 'a.N tree'). Due to its pronominal gender system, English is classified as having three genders, a group comprising only 26 languages (10.1%) in Corbett's sample. To have four genders is even rarer, found in no more than 12 (4.7%) languages. An example of a language with four genders is Lak, as summarized in Table 7.2 adapted from Corbett (1991: 25):

Table 7.2 The Lak (Nakh-Daghestanian (Lak-Dargwa): Russia) gender system.

GENDER	CRITERION	EXAMPLE	GLOSS
I	male rational	las	husband
II	female rational	ninu	mother
III	other animate	nic	bull
IV	residue	nex	river

The four genders in Lak are assigned according to the following principles: genders I and II comprise only humans and spiritual beings; gender III comprises non-rational animates (various animals and insects) as well as most inanimate objects; gender IV comprises some animates (such as, for example, spiders and dragonflies), some inanimate objects, most liquids, and abstract nouns (Corbett 1991: 25). This kind of system is common, though by no means exclusive to, Nakh-Daghestanian languages spoken in the Caucasus.

The last group in Corbett's sample, comprising 24 (or 9.3%) languages, are those languages with five or more genders. The majority of the languages in the sample with very many genders can be found in sub-Saharan Africa, as shown by the black dots in Map 7.1 above. An example of a language with five genders is Ju|'hoan, where the gender assignment is roughly as shown in Table 7.3.

An example of a language with very many genders indeed is Nkore-Kiga, with 17 different noun classes, assigned roughly on semantic criteria such as animacy, shape and abstractness (Taylor 1985: 124ff).

Languages vary with respect to how the gender distinctions pattern in the independent personal pronominal system. As mentioned above, it is most common cross-linguistically to not have any genders at all. This holds for the independent pronouns too, as evidenced in Siewierska's (2011b) sample, where as many as 254 of 378 languages (67.2%) lack gender distinctions. Of those languages that do have gender

Table 7.3 The Ju|'hoan (Khoisan (Northern Khoisan): Angola, Namibia, Botswana) gender system (Dickens n.y.: 12ff).

GENDER	CRITERION	EXAMPLE	GLOSS
I	humans	g‖àq	aunt
II	animals & non-Ju|'hoan humans	!xó |'Hún	elephant white person
III	plants & various inanimates	n|làng	raisin
IV	long things	n!ámá	road
V	body parts	n|áí	head

distinctions in their independent personal pronouns, the most common pattern, with 61 languages (16.1%), is that found in English, where only the third person singular pronoun has any gender distinctions. This is also found in Xokléng:

Table 7.4 The Xokléng (Macro-Ge (Ge-Kaingang): Brazil) independent pronoun in the nominative (Wiesemann 1986: 361).

PERSON	SINGULAR		PLURAL	
1	nũ	'I'	nã	'we'
2	mã	'you'	mã, mẽ	'you'
3M	ta	'he'	óg	'they'
3F	ði	'she'		

As the table above shows, the masculine/feminine distinction made in the third person singular is not made in the third person plural, where the form *óg* 'they' gives no information as to gender.

In 42 languages (11.1%) gender distinctions are restricted to the third person, but are found in both the singular and non-singular. This means that gender distinctions might be made in all numbers of the language, or in the singular and some other non-singular number. An example of the former is Worora (Australian (Wororan): Australia), which distinguishes between four genders (masculine, feminine, neuter-1 and neuter-2) in the third person singular, dual, trial as well as plural (Siewierska 2011b).

Some languages, 18 (4.8%) in Siewierska's sample, make gender distinctions in the third person as well as either the first or the second person. An example of a language with a maximum set of gender distinctions is Korana, where three genders (common, masculine and feminine) are distinguished in all persons and all numbers:

Table 7.5 The Korana (Khoisan (Central Khoisan): South Africa) independent pronouns (Siewierska 2011b citing Meinhof 1930: 43).

PERSON	GENDER	SINGULAR	DUAL	PLURAL
1. INCL	C	–	sam	sada
	F	–	sasam	sasē
	M	–	sakham	satjē
1. EXCL	C	–	sm	sida
	F	tita	sisam	sisē
	M	tire	sikham	sitjē
2.	C	–	sakhao	sadu
	F	sas	sasaro	sasao
	M	sats	sakharo	sakao
3.	C	ll'ãi'i	ll'ãikha	ll'ãinē
	F	ll'ãis	ll'ãisara	ll'ãidē
	M	ll'ãib	ll'ãikhara	ll'ãiku

Very rare systems are those that distinguish gender in the first or second persons, but not in the third, found only in two (0.5%) languages in Siewierska's sample: Burunge and Iraqw, both Afro-Asiatic Southern Cushitic languages in Tanzania. Angas, which is not in the sample, is another language with this kind of system:

Table 7.6 The Angas (Afro-Asiatic (West Chadic): Nigeria) independent pronouns (Burquest 1986: 80).

PERSON	GENDER	SINGULAR	PLURAL
1.	–	ŋán	mún
2.	M	ɣā	wún
	F	yī	
3.	–	nyə̄	mwá

As the table above shows, Angas makes two gender distinctions in the second person singular, but nowhere else.

One language (0.3%) in Siewierska's sample, Dagaare, has gender distinctions in the third person plural only, and nowhere else:

Table 7.7 The Dagaare (Niger-Congo (Gur): Ghana) independent pronouns (Siewierska 2011b citing Bodomo 1997: 71).

PERSON	GENDER	SINGULAR	PLURAL
1.	–	maa	tenee
2.	–	foo	yɛnee
3.	HUM	onɔ	bana
	NHUM		ana

As the table above shows, the Dagaare independent pronouns make no gender distinctions at all in the singular, while in the plural the two genders (human versus non-human) are distinguished in the third person.

The languages in APiCS display a rather similar pattern to those in the WALS, where the pattern is affected minimally by the subtraction of Ndyuka and Sango (none of which have gender distinctions). To not have any gender distinctions is by far the most common with 56 (of 73 or 76.7%) languages (Michaelis et al. 2013: feature 13). If a language has gender distinctions in the independent pronoun system, it will be in the third person. Most of the APiCS languages with gender distinctions in the independent pronoun system have that distinction in the singular only. This group comprises 12 (15.8%) languages. Notice that this proportion is almost exactly the same as that of the WALS sample with the same value (see Table 7.8 below). Four languages (5.3%), Kikongo-Kituba (Creole (Kikongo-Kimanyanga-lexified): DR Congo), Lingala (Creole (Bobangi-lexified): DR Congo),[108] Michif and Sri Lanka Portuguese (Creole (Portuguese-lexified): Sri Lanka) have gender distinctions in both singular and plural of the third person, while two languages (2.6%), Cape Verdean Creole of Brava and Cape Verdean Creole of Santiago (both Portuguese-lexified creoles of the Cape Verde Islands), have gender distinctions in the first and/or second person but not the third. The figures are summarized in Table 7.8.

Table 7.8 Figures for gender distinctions in independent pronouns in WALS (contact languages subtracted) and APiCS languages, absolute numbers in parentheses.

Value	WALS		APiCS	
No gender distinction	67%	(252)	76.3%	(58)
Gender distinction in 3SG only	16.2%	(61)	15.8%	(12)
Gender distinction in 3PL only	0.3%	(1)	–	
Gender distinction in 3SG&PL only	10.9%	(41)	5.3%	(4)
Gender distinction in 3 + 1 and/or 2	4.8%	(18)	–	
Gender distinction in 1 and/or 2 but not 3	0.5%	(2)	2.6%	(2)
		376		73

What the table above us tells us is that pidgin and creole languages are less likely than non-creole languages to have gender distinctions in the first or second persons of their independent pronouns, a phenomenon which is very rare in non-creoles too. Pidgin and creole languages are also somewhat more likely to lack gender distinctions in their independent pronouns than non-creoles, although in both samples this is the preferred strategy by far.

108. Note that this classification follows APiCS and differs from WALS, where Lingala is classified as a Niger-Congo Bantoid language.

7.1.2.2 *Classifiers*

Noun **classifiers** should be distinguished from noun classes/gender. As mentioned above, gender or noun classes are identified through agreement, which means that the forms of the words associated with the noun in question vary according to the gender of the noun. In other words, gender or noun class is identified through some kind of morphological process which has to involve some other word in the clause. Also, in gender or noun class systems, all nouns paradigmatically belong to one category or another. In other words, nouns obligatorily belong to one noun class or another in a closed grammatical system. Gender or noun class systems are also typically shown through various fusional processes, such as affixation, and as such tend to occur in languages that allow fusion. While some languages have a high number of genders or noun classes, the number is never unlimited. Classifiers, on the other hand, are free and invariant forms that assign nouns to a given category, usually semantically based. They are not dependent on agreement with a word outside themselves; "[t]hey are a type of non-agreeing noun categorization device, their choice being determined by lexical selection, and not by matching any inflectional properties of nouns with any other constituents of a noun phrase" (Aikhenvald 2000: 81). The same noun may take different classifiers, depending on context or what exactly the speaker wishes to express. For instance in Minangkabau the noun *limau* 'lemon' may take either the classifier *batang* 'CL:TREE' or the classifier *buah* 'CL:FRUIT':

Minangkabau (Austronesian (Malayic): Indonesia)

(99)

batang	limau	buah	limau
CL:TREE	lemon	CL:FRUIT	lemon
'lemon (the tree)'		'lemon (the fruit)'	

(Aikhenvald 2000: 84 citing Marnita 1996)

Classifiers may themselves function as nouns, and constitute a more open class of markers than gender or noun class markers. The choice of classifier is usually based on semantic characteristics in terms of everyday interaction. Languages may thus have classifiers labelling things as human, animals, plants, liquids, artefacts, or classifiers denoting various forms, shapes and structures, or classifiers denoting various kinds of social status, such as kinship, social functions, age, and so on.

There are other kinds of classifiers. Numeral classifiers "characterize nouns in numerical noun phrases and expressions of quantity" (Aikhenvald 2000: 426), usually according to animacy, physical properties (such as shape, size, structure, etc.), and so on.[109] Classifiers in possessive noun phrases may be either of various types. With

109. **Mensural numeral classifiers** are entities which provide non-countable nouns with a countable unit, such as *one **glass** of water* or *a **pound** of butter*. These kinds of classifiers basically give information on how a non-countable noun is to be measured. This kind of classifier is found in most languages (in

relational classifiers the relation between the possessor and possessed is categorized according to functional properties (whether the possessed item is to be eaten, sold, grown, etc.). With possessed classifiers, the possessed item is characterized in some way (in terms of animacy, physical properties, and so on). An extremely rare phenomenon in the world, so far known only in Dâw and possibly Hup (both Nadahup (Nadahup): Brazil), is the possessor classifiers, where the possessor is classified according to animacy (Aikhenvald 2000: 139 citing Martins 1994 and Moore & Franklin 1979). Locative classifiers and deictic classifiers are also quite rare. Locative classifiers categorize the head noun (according to physical properties and sometimes also animacy) in locative expressions. Deictic classifiers characterize the head noun (according to directionality and position in space, physical properties, and so on) in deictic expressions, i.e. with articles and demonstratives. Verbal classifiers appear on the verb but characterize the noun, usually either in subject or object position, according to various semantic properties (physical properties, position, and sometimes also animacy). For an accessible overview of classifiers, including numerous examples, see Aikhenvald (2000).

The most common type of classifier is the numeral classifier. In Gil's (2011) sample of 400 languages 140 (35%) have numeral classifiers. In 62 languages (15.5% of the entire sample) the numeral classifier is optional. An example of a language where numeral classifiers are optional is Persian, as shown in Example (100).

Persian (Indo-European (Iranian): Iran)

(100) a. do(-ta) pesær
two-CL:FOLD boy
'two boys'

b. bist(-jeld) ketab
twenty-CL:VOLUME book
'twenty books' (Mahootian 1997: 195)

In Example (100) the classifiers *-ta* 'fold' (used for any count noun) and *-jeld* 'volume' (used for books) are suffixed to the numeral if used. However, the use is not obligatory, as indicated by the parentheses.

In 78 (19.5%) of the languages in Gil's sample the numeral classifier is obligatory. This means that every time a noun is quantified with a numeral, there has to be a classifier in the numerical noun phrase. An example of a language where numeral classifiers are obligatory is Kambera. There are five classes, categorizing animacy and shape, as shown in (101):

varying degrees of grammaticalized states). **Sortal numeral classifiers**, on the other hand, are those classifiers that divide all nouns into semantic categories, irrespective of whether they are countable or not. What I term 'numeral classifier' in this section thus refers to sortal numeral classifiers.

Kambera (Austronesian (Central Malayo-Polynesian): Indonesia)

(101)			
	wua/mbua	round objects	
	pungu/mbungu	oblong objects	
	wàla/mbàla	flat, thin objects	
	iu/ngiu	animals	
	tau	humans	(Klamer 1998: 139)

When quantifying a noun one of these classifiers must be part of the numerical noun phrase. Compare the following:

Kambera (Austronesian (Central Malayo-Polynesian): Indonesia)

(102)					
	tailu	mbua	kajawa		
	three	CL:ROUND	papayas	'three papayas'	
	tailu	mbungu	pena		
	three	CL:OBLONG	pen	'three pens'	
	tailu	mbàla	kapambal		
	three	CL:FLAT&THIN	plank	'three planks'	
	tailu	ngiu	kamambi		
	three	CL:ANIMAL	goat	'three goats'	(Klamer 1998: 139)

In Example (102) the presence of a classifier is obligatory between the numeral and the noun that is being quantified. The choice of the classifier depends on its semantic characteristics, based on animacy and shape. Languages may have very large sets of numeral classifiers; Chontal Maya (Mayan (Mayan): Mexico), for example, has over a hundred (Suárez 1983: 88).

The languages in the APiCS sample exhibit much more radical figures than those in the WALS sample. While in both samples the majority of languages lack numeral classifiers, the figures differ in proportion. Recall that 35% of the languages in Gil's sample have numeral classifiers, most of which are spoken in South-East Asia. This figure changes only very insignificantly to 35.2% if we take out the two contact languages (Ndyuka and Chinese Pidgin Russian, both lacking classifiers) in the sample and thus have 398 languages in total. In the APiCS, however, only four of 76 languages (5.3%) have numeral classifiers: Ambon Malay (Creole (Malay-lexified): Indonesia), Chinese Pidgin English, Gullah (Creole (English-lexified): USA), and Sri Lanka Portuguese.[110] A full 94.7% lack them (Michaelis et al. 2013: feature 36). This means that pidgins and creoles seem less likely than non-creoles to have numeral classifiers.

110. Notice that two of these four languages are located in South East Asia.

7.1.3 Case

Case is a "system of marking dependent nouns for the type of relationship they bear to their heads" (Blake 2001:1). The term **head** refers to, very simplified, the central element in the construction, i.e. the element that establishes the function of the entire phrase. The head of the NP *that big fluffy blanket* is thus the noun *blanket*, since that is the central element. Without it the phrase would not make sense (cf. **that big fluffy*). The head usually governs the **dependent**, i.e. the element that is somehow attached to (dependent on) a head. The head of the noun is very often the verb of the clause or sentence, but it may also be other elements, such as adpositions or other nouns. This means that case is a grammatical way of overtly indicating what syntactic or semantic relationship the noun or noun phrase has to some other element of the clause or sentence within which it occurs. Traditionally case refers to various forms of morphological inflections on the NP, but it may also be argued that clitics and adpositions can function as case markers. For a very thorough overview of case and various types of case systems, see Blake (2001).

Languages differ as to how many and which kinds of cases they have. English has only one overtly marked case, the genitive. A typical, but not exclusive, feature of case systems is that they are paradigmatic. That means that the cases typically substitute each other; in other words, either case A is used, or case B, or case C but not a combination of more than one. The choice of which case to use depends on language-dependent syntactic and/or semantic rules. It would be entirely beyond the scope of this section to attempt to list the kinds of cases found throughout the world. However, some common core (or grammatical) cases will be mentioned, as these will then be particularly relevant for coming sections. It is very important here to keep in mind that the uses listed for the cases are wide generalizations. None of the cases listed are used exclusively for the functions given.

Many of us are familiar with the **nominative** case, which is typically used to mark the grammatical relation of the subject of the verb in the clause. Very simplified the subject can be thought of as the 'doer' of the action or the event. Examples of subjects are, for instance, *the man* in [*The man*] *danced* or *the girl* in [*The girl*] *bought* [*a book*]. The square brackets indicate which words form a single constituent (the unit that forms the relevant syntactic building block – in this case the subject and object). Notice that a constituent can consist of more than one word. The subjects in the sentences above are thus the NPs *the + man* and *the + girl*. The **accusative** case is typically used to mark the object in the clause, which, again very simplified, can be thought of as the 'goal' or 'receiver' of an event or action. Examples of an object are *a book* in the sentence above, or *the lamp* in [*The boy*] *switched on* [*the lamp*]. Sometimes two objects are needed, as in *The girl gave a book to her friend*. To distinguish between the two objects, the object that, very simplified, is the 'recipient' of the action or event is

labelled the indirect object. The **dative** case is often used to mark the indirect object. For more on constituents, see Chapter 10.

The **genitive** case is very often used to mark possession. Another common case is the **ergative**, which is typically used for the semantic role of agent. This is again very simplified indeed and will be discussed further in Chapter 9. Examples of an agent are *the girl* in *The girl gave a book to her friend*, or *the boy* in *The boy switched on the lamp*. The ergative case is typically used to mark the subject of a transitive clause.[111] Languages that have ergative case typically also have the **absolutive** case, which tends to be used for the semantic role of patient. An example of a patient is *the lamp* in the sentence above. Grammatical relations and semantic roles are not mutually exclusive: thus *the lamp* in the sentence above is both the object and the patient in the clause. This will be discussed further in Chapter 9. I stress again that the functions listed for the cases are very general and vary a great deal from language to language.

Other kinds of cases may involve notions of location, such as **locative** ('in/at X'), **adessive** ('near/by X') or **inessive** ('inside X'); or motion, such as **ablative** ('from X'), **allative** ('to X') or **illative** ('into X'); or states, such as **essive** ("at rest/be in/at", i.e. 'at/as X') or **abessive** ('without X'); or various other semantic roles, such as **benefactive** ('for X') or **comitative** ('with X').

It is common in the world, however, to not have any cases at all. In Iggesen's (2011) sample of 261 languages, 100 (38.3%) lack case altogether. This is the largest group in the sample. In these languages the various syntactic and semantic functions are expressed through other strategies, for example word order. English uses word order to distinguish between the grammatical relations of subject and object; in *The man saw the girl* the subject is *the man* and the object is *the girl*. If we swap the two NPs and get *The girl saw the man*, the subject changes and is now *the girl* while *the man* is the object. This is common among languages with small or no case systems. Due to the fact that English has a genitive case (the *'s* in, for example, the *dog's collar*), English is counted as having two cases in Iggesen's database because of the paradigmatic contrast between the genitive and the unmarked form. There are 32 (12.3%) languages in Iggesen's sample such as English, which are counted as having small case systems.[112]

---

111. A transitive clause is a clause where the verb demands two so-called arguments, which means that the verb needs two central elements for the clause to make sense. For example the verb *buy* is transitive in that someone has to buy something; a clause like **the man bought* is ungrammatical. The verb needs two arguments. For more on transitivity, see 9.1.3.2.

112. I have modified Iggesen's values somewhat, as follows: languages with small case systems (Iggesen's 'languages with 2 case categories' + 'languages with 3 case categories'), languages with medium-sized case systems (Iggesen's 'languages with 4 case categories' + 'languages with 5 case categories'). The three remaining values, languages with large case systems (6–7 cases), languages with very large cases systems (8–9 cases) and languages with extremely large case systems (10 cases

An example of a language listed as having three cases in Iggesen's sample is Modern Greek with nominative, accusative and genitive (Ruge 1984).[113] The smallest group, with 21 languages (8%), in Iggesen's sample is that with medium-sized case inventories of four or five cases. An example of a language with four cases is Kashmiri (Indo-European (Indic): India), with nominative, ergative, dative and ablative (Wali & Koul 1997: 151). Ngiyambaa (Australian (Pama-Nyungan): Australia) is an example of a language with five cases: absolutive, ergative/instrumental, dative, locative and circumstantive (Donaldson 1980: 82). Most common among the languages that have cases is to have a large inventory of six or seven cases; 37 (14.2%) languages of the sample fall into this group. Turkish is an example of a language with six cases in its case system: nominative, accusative, dative, locative, ablative and genitive (Kornfilt 2003: 212) and Kannada is an example of a language with seven cases: nominative, accusative, genitive, dative, locative, instrumental ('with X') and ablative (Sridhar 1990: 156). To have very large case inventories (eight or nine cases) is not quite as rare as one might think if one is mainly used to Western European languages: 23 (8.8%) languages display this pattern. West Greenlandic has eight cases (Fortescue 1984: 206) and Udihe has nine (Nikolaeva & Tolskaya 2001: 106), as shown in (103):

(103)

West Greenlandic (Eskimo-Aleut (Eskimo): Greenland)			**Udihe** (Altaic (Tungusic): Russia)	
absolutive	*-q/t/k/Ø*		nominative	*Ø*
ergative[114]	*-(u)p*		accusative	*-wA*
instrumental	*-mik*		dative	*-du*
allative	*-mut*		allative[115]	*-tigi*
locative	*-mi*		locative	*-lA*
ablative	*-mit*		prolative	*-li* ('along/through X')
prosecutive	*-kkut*	('through X')	ablative	*-digi*
equative	*-tut*	('like, as X')	instrumental	*-zi*
			destinative	*-nA* ('destined for X/destination')

It might come as a surprise that it actually is slightly more common to have extremely large case systems: 24 (9.2%) languages in Iggesen's sample have ten cases or more. An example of a language with a very rich case system is Lezgian (Nakh-Daghestanian

---

or more) are adopted unchanged from Iggesen. Notice that Iggesen includes clitics and adpositions as case markers provided that they show a certain degree of bondedness with the host noun.

113. It should be noted that Iggesen does not count non-syntactic cases such as the vocative (used to address someone), which in other interpretations would be the fourth case for Modern Greek.

114. Called 'relative' in Fortescue (1984).

115. Called 'lative' in Nikolaeva & Tolskaya (2001).

(Lezgic): Russia) with 18 cases, four core (or grammatical cases) – absolutive, ergative, dative and genitive – and 14 local cases (Haspelmath 1993: 74), as shown in (104). Notice that the ergative suffix is combined with all cases except the absolutive, which has no morphological marking. Notice also that the local cases divide neatly into five groups of localizations (ad- 'at', post- 'behind', sub- 'under', super- 'over/on', and in- 'in') of three locatives each (essive 'at rest', elative 'from' and directive 'towards'). The last logical combination of in + directive is lacking.

(104) **Lezgian** (Nakh-Daghestanian (Lezgic): Russia)

absolutive	*Ø*	
ergative	*-re*	
genitive	*-re-n*	
dative	*-re-z*	
adessive	*-re-w*	('at X')
adelative	*-re-w-aj*	('from X')
addirective	*-re-w-di*	('toward X')
postessive	*-re-q*^{h}	('behind X')
postelative	*-re-q*^{h}*-aj*	('from behind X')
postdirective	*-re-q*^{h}*-di*	('to behind X')
subessive	*-re-k*	('under X')
subelative	*-re-k-aj*	('from under X')
subdirective	*-re-k-di*	('to under X')
superessive	*-re-l*	('on X')
superelative	*-re-l-aj*	('off X')
superdirective	*-re-ldi*	('onto X')
inessive	*-re*	('in X')
inelative	*-räj*	('out of X')

Nakh-Daghestanian languages are famous for their large case systems, but as with Lezgian, the systems often form rather straightforward agglutinative combinations of a limited set of markers. This is also what we find in the northern dialects of Tabassaran (Nakh-Daghestanian (Lezgic): Russia), which are reputed to have extremely large case systems indeed, with as many as 53 cases (Comrie 1981c: 209, Comrie & Polinsky 1998: 96). However, here too we have a limited set of markers of four core cases (absolutive, ergative, genitive and dative) which may combine with eight locative or orientation markers and two motion markers (see Comrie & Polinsky 1998, also for a demonstration on how languages with a rather limited amount of cases can be claimed to have enormous case systems). Ugric languages are also famous for their rich case systems; in fact the language that Iggesen (2011) lists as having the most cases in his sample is Hungarian (Uralic (Ugric): Hungary), with 21 productive cases, although analyses for Hungarian vary radically and can range from 17 to 28

(see Spencer 2008 for references and, above all, for a discussion that Hungarian lacks a true case system, but that what look like case markers are actually 'fused postpositions' similar to the French portmanteau *du* = *de le* 'of the.M.SG'.)

The remaining languages in Iggesen's sample, 24 (9.2%), are analysed by Iggesen as having exclusively borderline case marking, by which he means that core (or grammatical) case is not marked, while such semantic notions as location or instrument are marked.

The vast majority of languages with case systems mark their cases through suffixation. In Dryer's (2011t) sample of 1032 languages, 653 (63.3%) have case marking. Of these, a full 452 (69.2% of the languages with case) mark cases through suffixation. West Greenlandic, Udihe and Lezgian are all examples of languages with suffixing case markers, as shown above. While case marking through prefixing is rare, it can be found in 38 languages (5.8% of the languages with case) in Dryer's sample. An example of a language with prefixing case markers is Gapapaiwa where *ku-* 'to' (LOCative) is prefixed on the noun:

Gapapaiwa (Austronesian (Oceanic): Papua New Guinea)

(105) ta-rui ku-vao
1INCL:PL-go.in to-garden
'We will go into the garden.' (McGuckin 2002: 314)

The second most common strategy in Dryer's sample, found in 123 languages (18.8% of the languages in the sample with case) is to have postpositional clitics – recall that clitics are units that are phonetically bound, but that are syntactically independent of their host. An example of a postpositional case marker is *-ṛa* 'LOCative' in Ungarinjin:

Ungarinjin (Australian (Wororan): Australia)

(106) [dambun ŋininga]=ṛa
camp my=LOC
'at my camp' (Rumsey 1978: 85)

In Example (106) the locative case marker is placed at the end of the whole NP, even though the last word of the NP is not the noun. A few languages in Dryer's sample – 18 (2.8% of the languages with case) to be precise – use prepositional clitics to mark case, such as Cayuvava.

Cayuvava (Isolate: Bolivia)

(107) ji=[ka'reeča datï]
OBL=other place
'in another place' (Dryer 2011t citing Key 1967: 51)

In (107) the oblique case marker *ji-* cliticizes to the first word of the NP, even if that is not the noun. Extremely rare is to have case marking inpositional clitics, which is found in only seven languages (1.1% of the languages with case) in Dryer's sample. An example of such a language is Anindilyakwa:

Anindilyakwa (Australian (Anindilyakwa): Australia)

(108) ampwarriya [arimwa=mwantja aka][116]
2SG.sit 3:4.big=LOC 3:4.tree
'Sit next to the big tree!' (Leeding 1989: 308)

In Example (108) the locative case marker *-mwantja* cliticizes onto the first word of the noun phrase, in this case the adjective. Notice that this kind of construction was listed as an inposition in Dryer (2011i); cf. Example (72) above), showing that the boundary between case and adpositions is a fuzzy one.

Another extremely rare strategy is to mark case through tone; this is found in only five languages (0.8% of the languages with case) in Dryer's sample: Jamsay (Niger-Congo (Dogon): Mali), Maasai, Maba (Nilo-Saharan (Maban): Chad), Nandi and Shilluk (Nilo-Saharan (Nilotic): Sudan). The Agar dialect of Dinka, which is not in Dryer's sample, also marks case through tone only.

Agar Dinka (Nilo-Saharan (Nilotic): Sudan)

(109) a. bɛ́ŋ à-tʊ̀oc dɛ̰̀ɔk
chief.ABS DECL-send boy
'The chief is sending the boy.'
b. dɛ̰̀ɔk à-tʊ́oc bɛ̀ŋ
boy DECL-send.PASS chief.OBL
'The boy is being sent by the chief.' (Andersen 2002: 7)

In Example (109) above the only difference between 'chief' in the absolutive and the oblique case is the tone: *bɛ́ŋ* versus *bɛ̀ŋ*. Even more rare is to mark case through stem change in the noun, a strategy found in only one language (0.2% of the languages with case) in Dryer's sample, Nuer. Example (110) from Nuer shows how stem change is the only way of differentiating between nominative and genitive case.

Nuer (Nilo-Saharan (Nilotic): Ethiopia, Sudan)

(110) 'cow' NOM.SG yaŋ 'drum' NOM.SG bul
GEN.SG yaaŋ GEN.SG buɔ̈ɔ̈l (Wright 1999: 85)

116. The notation 3:4 indicates noun class. The 3:4 noun class in Anindilyakwa categorizes "invisible items and items with a lustrous appearance (animate and inanimate)" (Leeding 1989: 229).

In the example above the only difference between the cases is in the change of stem through either mere lengthening (*yaŋ* versus *yaaŋ*) or through diphthongation (*bul* versus *buɔ̯ɔ̯l*).

Finally, eight languages (1.3% of the languages with case) in Dryer's sample exhibit a mixed strategy, combining at least two of the suffixing, prefixing, tonal or stem change methods for coding case.

7.2 NP syntax

We have seen that NPs may consist minimally of a noun or a substitute for a noun (such as a pronoun), but that they may also consist of several words that together form a constituent unit. In the latter case there is typically a core or main element, a head, which the other elements in the phrase, the dependent(s), relate to one way or another. We have also seen that the elements in an NP may be marked for various categories, such as number, gender or class, and case. Several words that belong together in an NP may be overtly marked to show that they belong together; they agree with each other formally.[117] The whole NP can in turn be a dependent to a head outside the NP, such as a verb or a preposition. These heads outside the NP may determine which form the NP requires; they govern the NP.

7.2.1 Heads and dependents

When several elements combine to form a unit, there is typically some kind of syntactic relation between them in the sense that there is an organization by which one element forms the core of the unit which the other units relate to. For instance in the NP unit *John's book*, the core element of the phrase, the **head**, is *book*, while the **dependent** element is *John's*, since the NP *John's book* refers to a specific book and not a specific John. In the NP *red flower* the head of the phrase is *flower* and the dependent is *red*, since we are talking about a kind of flower (a red one) and not a kind of red (a flowery one). This is also true for longer phrases, such as *the king of Sweden's crown*, where *crown* is the head and *the king of Sweden* is the dependent (we are talking about a specific crown and not a specific king of Sweden). A phrase can also have an adposition as its head, with the various objects of the adposition being the dependents. These kinds of phrases are, unsurprisingly, called adpositional phrases – in English we have prepositional phrases (PPs), since the adpositions of English are prepositions. An example of an adpositional phrase is *from the red house*, where the preposition *from* is

117. While the controller of agreement is usually a nominal, other elements may also show agreement, such as elements in a VP showing agreement for tense, mood and/or aspect.

the head and the NP *the red house* is the dependent (which then, in turn, has its own inner syntax of heads and dependents).[118]

Languages differ as to whether they overtly mark this syntax or not. The less morphological marking a language makes use of, the more likely it will be that head and dependent syntax will be unmarked. Notice, however, that languages may mark some kinds of relationship morphologically, while others remain unmarked. English is an example of this, where possessive phrases are marked with the genitive *'s* (e.g. *John's book*), while attributive phrases and adpositional phrases are unmarked (e.g. *red flower* and *from the house* respectively). Languages also differ as to whether the relationship is marked on the head or on the dependent or both.[119]

English is an example of a **dependent marking** language in possessive phrases, where the possessor and not the possessed noun carries the marker: *John* in *John's book* is the possessor while *book* is the possessed noun. An example of a **head marking** language is Fijian:

Fijian (Austronesian (Oceanic): Fiji)

(111) a liga-i Jone
ART hand-POSS PN
'John's hand' (Dixon 1988: 120)

In (111) the possessive marker *-i* suffixes to the possessed noun (the head) *liga* 'hand' and not the possessor (the dependent) *Jone* 'John'. An example of a **double marking** language, i.e. a language that marks both the head and the dependent, is Turkish:

Turkish (Altaic (Turkic): Turkey)

(112) Ayşe-nin araba-sı
PN-GEN car-3SG
'Ayşe's car' (lit. 'Ayşe's his-car') (Kornfilt 1990: 633)

---

118. The binary opposition of head and dependent carries over to other and larger units too. For instance, the inner organization of verb phrases (VPs) is that the auxiliary verb is the head and the lexical (or main) verb is the dependent (see Chapter 8 for more on VPs). On a higher level, the predicate (usually the verb) is the head of the clause, while the arguments and adjuncts are the dependents (see further Chapters 9 and 10). Likewise, the main-clause predicate is the head of the sentence while the subordinate clauses are the dependents (see further Chapter 11).

119. There are various other possibilities too, one example being free (or floating) marking, where the marker always has a certain position in the phrase irrespective of where the head or dependent are located (for instance, a linker clitic that always attaches to the first word of the phrase, irrespective of what word that is, the so-called Wackernagel position.).

In (112) both the possessed noun (the head) *araba* 'car' and the possessor (the dependent) *Ayşe* carry overt marking. Indonesian is an example of a language with no marking at all in possessive phrases, but where simple **juxtaposition** (that the elements are placed next to each other) suffices, as Example (113) shows.

Indonesian (Austronesian (Sundic): Indonesia)

(113) rumah Tomo
house PN
'Tomo's house' (Sneddon 1996: 144)

In their sample of 236 languages, Nichols & Bickel (2011a) found that 98 (41.5%) of the languages were dependent marking, like English, while 78 (33.1%) languages were head marking, like Fijian, and 22 (9.3%) were double marking, like Turkish. A few languages, six to be precise (2.5%), had other kinds of strategies while 32 (13.6%) lacked any kind of marking and simply juxtaposed the two elements.

As mentioned, attributive phrases, containing a noun and one or several modifying adjectives, may also be either head or dependent marking. An example of a dependent marking language in attributive phrases is Chechen, where an agreement prefix is placed on the adjective (the dependent) to show agreement with the covert gender class noun (the head):

Chechen (Nakh-Daghestanian (Nakh): Russia)

(114) a. d-ovxa xi
AGR-hot water
'hot water'

b. j-ovxa šura
AGR-hot milk
'hot milk' (Nichols 1986: 61)

In Example (114a) the agreement prefix *d-* indicates the covert gender class of the noun *xi* 'water' and is placed on the adjective (the dependent) *ovxa* 'hot'; the noun itself (the head) remains unmarked. With a 'covert gender class' I mean that the gender is not overtly expressed, even though it is part of the inherent semantics of the noun, as the adjective agreement shows. When the adjective modifies a noun of a different gender class, as in (114b), the prefix changes accordingly (in this case to *j-*).

An example of a head marking language in attributive phrases is Shuswap, where the marker for oblique case (called the relative case for this language) is placed on the noun (the head) and not the adjective (the dependent), as shown in Example (115).

Shuswap (Salishan (Interior Salish): Canada)

(115) wist t-citx
high REL-house
'high house' (Nichols 1986: 61)

In the example above the adjective (the dependent) *wist* 'high' is unmarked while the noun (the head) *citx* 'house' carries a relative (oblique) case marker.

Adpositional phrases may also be either head or dependent marking. An example of a dependent marking language for adpositional phrases is German, where the preposition *wegen* 'because.of' requires the genitive case:

German (Indo-European (Germanic): Germany)

(116) wegen Regen-s wurde das Spiel unterbrochen
because.of rain-GEN AUX.PAST DEF game suspend.PTCPL
'Because of rain the game was suspended.' (source: personal knowledge)

In the adpositional phrase in (116), *wegen Regens* 'because.of rain', the marker is on the noun (the dependent) *Regen* 'rain' while the preposition (the head) *wegen* 'because.of' remains unmarked.[120] An example of a head marking language in adpositional phrases is Tzutujil:

Tzutujil (Mayan (Mayan): Guatemala)

(117) ruu-majk jar aachi
3.SG-because.of DEF man
'because of the man' (lit. 'he-because.of the man') (Dayley 1981: 216)

In (117) above, the preposition (the head) *majk* 'because.of' takes an agreement marker with the noun (the dependent) *aachi* 'man', while the noun (the dependent) remains unmarked.

7.2.2 Government and agreement

We have seen that several elements can be linked to form one syntactic unit, and that there may be some kind of binary organization to these elements in that one is the head and the other(s) is (or are) the dependent(s). In morphologically complex languages this linkage can be one of **agreement** (also termed **concord**), in which case the dependents take a morphological shape that corresponds to that of the head (or, more accurately, the controller determines the form of the target, see below). In other words, if the head has the inherent semantics of 'male + singular + nominative' then the dependents must also take the form 'male + singular + nominative' in order to show that they form a unit with the head. Similarly, if the head has the inherent semantics of 'female + plural + genitive' then the dependents must also take the shape

120. This example represents rather formal language. More common would be to include the article in the construction, as in *wegen des Regens* (because.of ART.M.SG.GEN rain-GEN).

that signals 'female + plural + genitive'. The crucial thing about agreement is that the dependents vary in shape systematically according to the semantics of the head. The linkage between elements in a unit can alternatively be one of **government**, in which case the head determines the morphological shape of the dependents (or, more accurately, where the governor determines the shape of the governee, see below), but where the shape of the dependents does not give any information about any inherent semantics of the head. For example, an adposition might govern the genitive case, as in Example (116) above, but the fact that the dependents are in the genitive gives no particular information about the inherent semantics of the adposition.

Patterns of agreement vary greatly cross-linguistically and it would be beyond the scope of this section to give anything more than a brief overview of the most prototypical examples. For a very thorough introduction to agreement, see Corbett (2006). The element that triggers agreement, the **controller**, is typically a nominal, while the elements whose form is determined by the controller, the **targets**, may be of different sorts but are prototypically adjectives or verbs. The features involved in agreement are typically gender, number and person. Consider the Italian example below:

Italian (Indo-European (Romance): Italy)

(118)

a.	il	libr-o	piccol-o
	DEF.M.SG	book-M.SG	small-M.SG
	'the small book'		
b.	i	libr-i	piccol-i
	DEF.M.PL	book-M.PL	small-M.PL
	'the small books'		
c.	la	cas-a	piccol-a
	DEF.F.SG	house-F.SG	small-F.SG
	'the small house'		
d.	le	cas-e	piccol-e
	DEF.F.PL	house-F.PL	small-F.PL
	'the small houses'		

(source: personal knowledge)

In Example (118) the definite article as well as the adjective, the targets, systematically change shape according to the noun, the controller. If the controller is a masculine noun, as *libro* 'book(M)' is, the targets have to be in a masculine form and if the controller is a feminine noun, as *casa* 'house(F)' is, the targets have to be in a feminine form. If the controller is in the singular, then the targets are also in the singular, and if the controller is in the plural, then so are the targets.

Example (118) shows a typical target type, the adjective. Another typical target within the clause (as opposed to within the phrase) is the verb, which often agrees in person and number with the subject, as is the case in French.

French (Indo-European (Romance): France)

(119) a. Jean aim-e les chat-s
PN love-3SG ART cat-PL
'Jean loves cats.'

b. nous aim-ons Jean
1PL love-1PL PN
'We love Jean.' (source: personal knowledge)

In Example (119) the target, the verb *aimer* 'love', is inflected for the person and number of the subject: if the subject is in the third person singular, this is signalled on the verb through a third person singular affix. If the subject changes in person or number, then the affix must also change.

Other targets may be pronouns, numerals, adverbs, adpositions, complementizers, coordinating conjunctions, particles and other nouns. For examples and discussion, see Corbett (2006).

While with agreement the feature specification of the targets are determined by the feature specifications of the controllers, with government the feature specification of the governed element (the **governee**) is determined merely by the presence of the governing element (the **governor**): "the governing member [i.e. the governor] imposes specific restrictions on the morphosyntactic properties of the governed member [i.e. the governee], but does so without (necessarily) sharing any of its properties" (Stump 1998: 24). In other words, the governor does not (necessarily) have the inherent semantic content that the shape of the governee indicates. While nominals are the typical controllers for agreement and targets can vary, the opposite is true for government: the governor can be of different kinds but the governee is typically a nominal. The feature involved in government is typically case.

We have seen that adpositions may govern case (the preposition *wegen* 'because. of' governing the genitive in Example (116) above). This is also the true for Khwarshi, where postpositions govern different cases in the NP. For example *žoqu'uža* 'behind' governs the genitive-2 case, *gił* 'under' governs the interessive or subessive case and *roq'ihol* 'according to' governs the subessive case, as shown in Example (120):

Khwarshi (Nakh-Daghestanian (Avar-Andic-Tsezic): Russia)

(120) a. aⁿc-ma-la žoqu'uža
door-OBL-GEN2 behind
'behind the door'

b. łe-ł gił
water.OBL-INTER under
'in the water'

c. q'ut'i-λ'o roq'ihol
deal-SUB according.to
'according to the deal' (Khalilova 2009: 134–5, 140)

GOVERNMENT OF AGREEMENT

In the Germanic languages two paradigms for adjective inflection were developed, the 'strong' one and the 'weak' one. This distinction has been lost in English due to the loss of inflectional morphology, but can still be seen clearly in German (Indo-European (Germanic): Germany). It is also still present in the Scandinavian languages. In the strong paradigm we have the forms distinguishing between singular and plural number, four cases and three genders (in the singular only), shown in Table 7.9. Notice that the strong form paradigm has 16 cells but only five distinct forms. The weak paradigm has the forms shown in Table 7.10.

Table 7.9 The German strong adjective

	SINGULAR			PLURAL
	MASCULINE	FEMININE	NEUTER	
NOMINATIVE	rot-er	rot-e	rot-es	rot-e
ACCUSATIVE	rot-en	rot-e	rot-es	rot-e
DATIVE	rot-em	rot-er	rot-em	rot-en
GENITIVE	rot-en	rot-er	rot-en	rot-er

Table 7.10 The German weak adjective

	SINGULAR			PLURAL
	MASCULINE	FEMININE	NEUTER	
NOMINATIVE	rot-e	rot-e	rot-e	rot-en
ACCUSATIVE	rot-en	rot-e	rot-e	rot-en
DATIVE	rot-en	rot-en	rot-en	rot-en
GENITIVE	rot-en	rot-en	rot-en	rot-en

The five distinct forms have been reduced to two in Table 7.10. In German there is also a mixed paradigm which shares some endings from the strong paradigm and some from the weak, with a total of four distinct forms. The adjective must agree with the noun according to one of these paradigms. The choice of paradigm is governed by the presence and kind of determiner. If there is no determiner, the strong paradigm is used; however, as soon as a determiner carries any case marking, the weak paradigm is used (the mixed paradigm is used with those determiners that vary with respect to overt case marking). In other words, we have agreement with adjectives, but the choice of paradigm for that agreement is an instance of government.

Verbs also commonly govern case. In Modern Greek, for example, most direct objects are in the accusative case, but with some verbs, such as *ftano* 'suffice', the direct object is in the genitive.

Modern Greek (Indo-European (Greek): Greece)

(121) tis gynaikas ftanei to psomi
DEF.F.SG.GEN woman.SG.GEN suffice.3SG DEF.M.SG.NOM bread
'The bread is enough for the woman.' (adapted from Ruge 1984:107)

In Example (121) the subject is *psomi* 'bread' and is in the nominative, as would be expected. However, the object, *tis gynaikas* 'the woman', is not in the accusative, which is the usual case for expressing direct objects in Modern Greek, but in the genitive. The noun phrase 'the woman', which constitutes the object here, the governee, is in the genitive due to the verb, the governor. In Russian the numerals 'two', 'three' and 'four' govern the genitive case and singular number:

Russian (Indo-European (Slavic): Russia)

(122) dva/tri/četyre žurnal-a
two/three/four magazine-SG.GEN
'two/three/four magazines' (adapted from Corbett 1993: 13)

In (122) the noun, the governee, is governed by the numeral, the governor, to take the genitive case and singular number. The numeral itself is in the nominative.[121]

7.3 Nominal categories and syntax in sign languages

As with spoken languages, signed languages have means for expressing more of the same. Nouns may either be unmarked for number, transnumeral (or have general number, see above), or they may contrast between singular and plural, in which case the plural is typically marked through reduplication. For instance, the plural form of DGS HOUSE is to repeat the sign three times (simple reduplication) while the plural for PERSON is to repeat the sign sidewards three times (sideward reduplication), as shown in Figure 7.2 below.[122]

Not all nouns can be reduplicated, in which case they remain unmarked for number. It seems common for sign languages to have phonological restrictions on which nouns can take reduplication, although the restrictions vary across languages. In DGS, for example, body-anchored signs (where the place of articulation involves a part of the body, such as the head or the non-dominant arm or hand) and signs with complex movements cannot be reduplicated. It would be ungrammatical to repeat the sign for DOCTOR or CAR in Figure 7.3 in an effort to express pluralization.

---

121. In actuality the situation is much more complex: the numeral needs to agree with the noun for gender, but since Russian only distinguishes for gender in the singular, the noun has to be in the singular. So while the numeral governs the case and number for the noun, and as such is the governor of the noun, the noun in turn controls the gender of the numeral, which makes the numeral a target for the noun. This is termed 'collaborative agreement' by Corbett (2006: 85). For more on the complexities of Russian numerals, see Corbett (1993).

122. Strictly speaking this is actually triplication, which is perceptually less ambiguous for the non-signer.

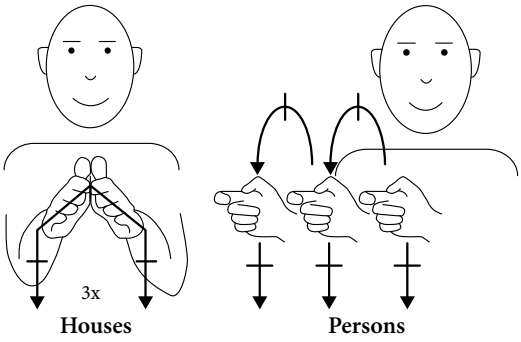


Figure 7.2 Example of a DGS (Sign Language: Germany) plural for HOUSE and PERSON (source: personal knowledge). Illustration: Maria Johanson. Used with permission.

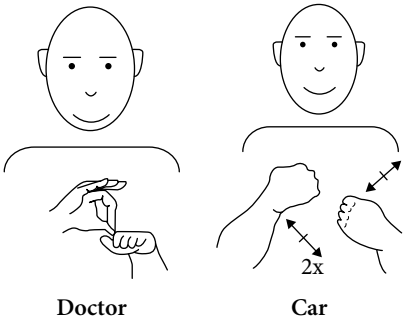


Figure 7.3 Example of a DGS (Sign Language: Germany) body-anchored sign (DOCTOR) and a complex movement sign (CAR) (source: personal knowledge). Illustration: Maria Johanson. Used with permission.

It is also common for sign languages to restrict plural marking to NPs lacking any other kind of quantifier. For instance something like *MANY/TWO HOUSE++[123], where the quantifier MANY or the numeral TWO are part of the NP, would not be grammatical in DGS. In LIS, on the other hand, the co-occurrence of a quantifier does not block plural inflection (Pizzuto & Corazza 1996: 184). Some sign languages lack plural inflection for nouns altogether. In ST, for example, nouns remain unmarked for number (Ahlgren & Bergman 2006: 29).

Many sign languages have more number values in their pronominal systems. The dual, for instance, is common across sign languages, and some, like Jordanian Sign

123. '+' indicates each reduplication of the sign. '++' thus means that the sign has been performed three times (once for the gloss and once for each '+').

Language (LIU: Jordan) may even specify for three (trial), four (quadral) and five (quintuple). This is done indexically, where the number of fingers extended specifies the number of referents meant.

Sign languages do not tend to have grammatical gender or noun classes. However, as we have seen (5.3), classifiers are very common indeed. As mentioned, these classifiers modify the verb they occur with and are better thought of as 'verbal classifiers' and not noun classifiers, as they do not categorize a type of noun, but specify semantic properties of the argument(s).

Case marking is also a minor phenomenon in sign languages. One exception is the genitive suffix in Auslan, as in MOTHER-*gen* SISTER-*gen* HUSBAND 'mother's sister's husband' (Johnston 2006: 325). This suffix, however, is only limited to a few lexical items and is not obligatory. To express possession simply by juxtaposition (MOTHER SISTER instead of MOTHER-*gen* SISTER), as is common in most sign languages, is equally grammatical (Johnston & Schembri 2009: 140).

Juxtaposition of elements that belong together in a phrasal unit is a common strategy for sign languages, especially for NPs; inflectional morphology is predominantly found in the VP. Thus, as mentioned, possession is commonly expressed by simply placing the two elements next to each other. While this is also found in spoken languages, it is not as common as in signed languages (see 7.2.1 above).

Due to the fact that most inflectional morphology is found in the VP and not the NP in sign languages, there is relatively little agreement in NPs. In DGS, for example, if an NP contains a noun and a modifying adjective, plural marking will be indicated on the noun only and not the adjective, which, incidentally, makes DGS a head-marking language. Thus BOOK++ GOOD 'good books' is grammatical while *BOOK++ GOOD++ is not. This is also true for ASL and ISL (Pfau & Steinbach 2006b: 171). However, in other sign languages, such as NGT, ÖGS, LIS and Hausa Sign Language (HSL: Nigeria), number agreement on the adjective is an optional possibility (Pfau & Steinbach 2006b: 171 with further references). Notice that the plural marking on nouns is often blocked if a numeral or other kind of quantifier is part of the NP. In other words, plurality is often only expressed once in an NP, either through a quantifier or through (usually optional) plural marking on the noun. In LIU, for example, the NP 'three deaf boys' is signed BOY DEAF THREE (Hendriks 2008: 69).

7.4 Summary

Nouns and noun phrases tend to be subcategorized according to various criteria. Noun phrases may be coded for number, obligatorily or optionally. Number values vary, but the contrast singular versus plural is most commonly found. The associative plural does not indicate number for an entity (usually a human), but that someone else is attached to that entity.

Noun phrases are often categorized into different genders or noun classes, usually according to a system based on semantic criteria. Languages vary radically as to how many genders they have, from none at all to up to 20. Classifiers should not be confused with noun class, as they are free and invariant forms that assign nouns to a given noun class.

Nouns phrases may have an overt marking for the syntactic relationship between constituents: case markers, found in a great deal of languages in the world. Languages vary as to how many cases they have, ranging from no cases at all to more than 20.

Noun phrases containing several elements usually have an internal syntax where one element is the head, while the other elements are the dependents of the head. Languages vary as to whether the various markers are coded on the head or the dependent(s). Agreement is when the dependents vary in shape systematically according to the semantics of the head. Government is when the mere presence of the head determines a certain shape of the dependent, irrespective of the semantics of the head.

Sign languages may also mark their noun phrases for number, typically plural for nouns but more values for pronouns. Plural is typically marked though reduplication and is not necessarily obligatory. Gender or noun class and case play very minor roles in sign languages. Classifiers are prolific, but belong to the verb phrase rather than the noun phrase. Sign languages vary with respect to agreement in attributive phrases.

7.5 Keywords

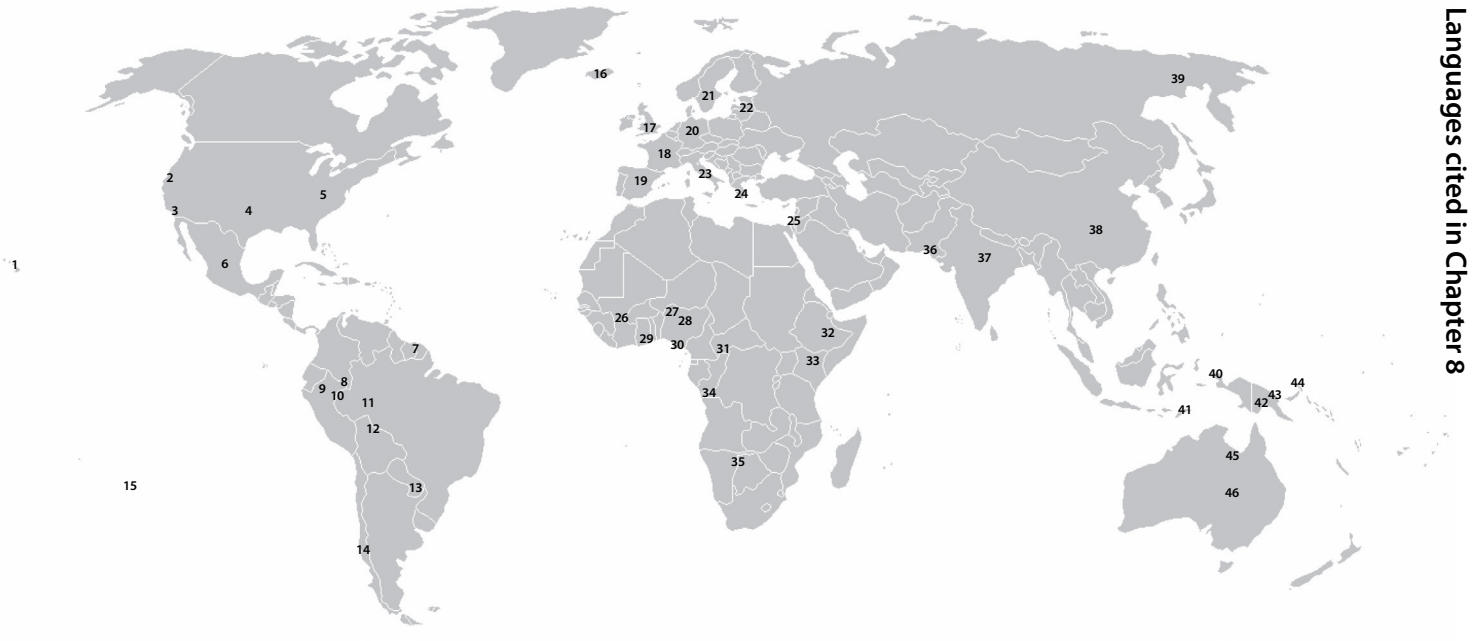
agreement
alienable/inalienable possession
animacy
associative plural
case
classifier
dependent
government
head
head/dependent marking
noun class/gender
number

7.6 Exercises

1. What is the **Animacy Hierarchy**?
2. How do pidgin and creole languages pattern with respect to nominal plural marking compared to non-creoles?
3. What is the difference between **government** and **agreement**?
4. How is possession indicated in sign languages and how does that compare to spoken languages?
5. Is the following statement true or false? Motivate your answer.

   There is no difference between gender and noun classifiers.

Languages cited in Chapter 8



1 Hawai'i Creole English
2 Wintu
3 Cupeño
4 Comanche
5 American Sign Language
6 Mexican Sign Language
7 Ndyuka
8 Yagua
9 Chamicuro
10 Matsés
11 Apurinã
12 Chácobo
13 Guaraní
14 Mapudungun
15 Rapanui
16 Icelandic
17 British Sign Language
English
18 French
19 Spanish
20 German
German Sign Language
21 Swedish
Swedish Sign Language
22 Latvian
23 Italian
24 Greek
25 Israeli Sign Language
26 Supyire
27 Hausa
28 Kuche
29 Ghanaian Pidgin English
30 Engenni
31 Sango
32 Oromo, Harar
33 Rendille
34 Kongo
35 Ju|'hoan
36 Urdu
37 Hindi
Indo-Pakistani Sign Language
38 Qiang
39 Yukaghir, Kolyma
40 Maybrat
41 Tugun
42 Foe
43 Anjam
44 Nalik
45 Kayardild
46 Australian Sign Language

Chapter 8

Verbal categories

Some of the most important verbal categories are tense, aspect and mood. Tense locates an event on a timeline, while aspect specifies the perspective taken on an event. Mood and modality code the attitude of the speaker towards a given proposition. While there are many other verb phrase operations to be found in and across languages, this chapter deals specifically with tense, mood and aspect. In 8.2 I give an overview of the major categories of tense, contrasting absolute versus relative tense, also mentioning remoteness. Section 8.3 gives a brief sketch of the perfect. In 8.4 I discuss aspect and its various subcategories, and contrast it with *Aktionsart* or actionality. Section 8.5 brings up mood and modality, where I first contrast realis with irrealis mood, then mention some different kinds of modalities. Section 8.6 gives an overview of the more common tense, aspect and mood or modality systems found in sign languages.

8.1 A very brief introductory note on the VP

Just as the noun phrase can consist of several elements that form a unit, so can the verb phrase (VP). Minimally the verb phrase consists of one word, a verb. Thus the sentence *The kangaroo jumped* contains one noun phrase (*The kangaroo*) and one verb phrase (*jumped*). However, a sentence like *The kangaroo should have been jumping* also only contains one noun phrase (*The kangaroo*) and one verb phrase (*should have been jumping*), even though the latter consists of several elements. Furthermore, in both sentences *The kangaroo jumped* and *The kangaroo should have been jumping* the verb phrases contains the same basic semantic content, that of the action JUMP. This is because both verb phrases contain the same lexical verb, or 'main' verb. It is the lexical verb that carries the semantic content of the verb phrase, the basic meaning of the event or action. The remaining elements in the verb phrase (*should have been*) in the sentence above are auxiliary verbs (cf. 6.2.2.5). These tend to be semantically empty and mainly convey grammatical or functional information.[124]

---

124. Later we will see that a verb phrase may be argued to contain yet other elements, namely nominal constituents (cf. Chapter 10).

Some very important verbal grammatical categories are tense, mood and aspect (TMA), all of which are interrelated one way or another.[125] Tense and aspect are both categories that deal with time in various ways, while mood deals with assertions about propositions. The traditional definition of the differences between the three is succinctly put in Chung & Timberlake (1985: 202):

> Tense, aspect, and mood are all categories that further specify or characterize the basic predication, which can be referred to as the event. Tense locates the event in time. Aspect characterizes the internal temporal structure of the event. Mood describes the actuality of the event in terms such as possibility, necessity, or desirability.

While these categories per definition blend into each other, as shall be seen, I will deal with them in turn. I stress, however, that this should not be taken to mean that these categories are clearly defined against each other. On the contrary, it is very common that grammatical markers denote a combination of two or more of the TMA categories. It would be beyond the scope of this chapter discuss the full range of complexities involved with TMA. The following sections will only highlight the main defining points for each category. For a very thorough discussion on various TMA categories and especially their origins, see Bybee et al. (1994). For a recent and accessible article-length overview on the typology of TMA, see de Haan (2010).

8.2 Tense

Tense is, very simplistically, a grammatical way for placing an event at a particular point in time. In other words, tense is the linguistic device used to indicate when an event took place. It is important to note here that tense refers to a grammatical category of markers used (often obligatorily) to locate an event on a timeline, and not to lexical items and expressions (such as *yesterday* or *at five o'clock*), which tend to be optional and context dependent. For instance, the suffix *-ed* in *Yesterday I walk**ed** home* is a past tense marker, whereas *yesterday* is not, even though they both can be said to place an event in the past on a timeline. However, the past marker *-ed* is obligatory; **yesterday I walk home* is not grammatical in (Standard) English, while a sentence

125. TMA is actually not restricted to the verb and verb phrases only, but can also be found as an inflectional category for nominals and other NP elements, as shown in Nordlinger & Sadler (2004 and subsequent). Of the 22 languages mentioned in Nordlinger & Sadler, most can be found in South America, but a handful are spread over sub-Saharan Africa, Australia and northern USA. The tense dependent demonstratives of Chamicuro mentioned below are an example of nominal tense. For further details, examples and references, see Nordlinger & Sadler (2004), but cf. also Tonhauser (2007) for a discussion on Guaraní specifically, challenging the interpretation of a nominal tense category there.

like *I walked home*, without the adverbial *yesterday*, is grammatically acceptable. For thorough discussions on tense, see Comrie (1985) and Dahl (1985), both of which are classics by now.

There are two conceptually different ways of expressing when something happened: to either relate the event to a given reference point (relative tense), or to relate it to the moment of speech (absolute tense). Each of these strategies will be discussed below.

If we imagine a timeline on which to place an event, a crucial concept is the present moment, the *Now*, as in Figure 8.1.

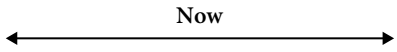


Figure 8.1 Timeline specifying only *Now*.

It is important to keep in mind here that the concept of *Now* lies with the narrator or the individual performing the speech act. Thus we often have the option to have a fictive *Now*, where we, in identifying with a character of a story, transpose a *Now* to something imagined. A minimal tense contrast would then be *Now* versus *not-Now*. This kind of system, which in essence lumps together the past and the future into one category, is as yet not known in any language of the world.[126] However, there are languages with a bipartite tense distinction, either before-*Now* versus not-before-*Now* (**past** versus **nonpast**) or after-*Now* versus not-after-*Now* (**future** versus **nonfuture**). An example of a past/nonpast system can be found in Harar Oromo:

Harar Oromo (Afro-Asiatic (Eastern Cushitic): Ethiopia)

(123) a. inníi magaláa deem-e
3SG.M market go-PAST
'He went to the market.'

b. inníi magaláa deem-a
3SG.M market go-NPST
'He goes/is going/will go to the market.' (Owens 1985: 83)[127]

While the sentence in Example (123a) is unambiguously in the past, the suffix *-a* in (123b) does not specify whether the person is going to the market now or will go later, but only that we are not referring to the past. An example of a language with a future/nonfuture system is Kolyma Yukaghir:

---

126. Yagua, however, has seven tense categories, two of which essentially do lump together the past and the future: PROXimate-1 and PROXimate-2 (Payne 1985a). However, they are not the only tenses of the language. See further below.

127. The suffix *-a* in (123b) is glossed 'imperfect' in the source.

Kolyma Yukaghir (Yukaghir (Yukaghir): Russia)

(124) a. terikie-die iŋd'e-t modo-j
old.woman-DIM sew-SS.IPFV sit-INTR.3SG
'The old woman is/was sitting and sewing.'

b. met qanin+ere kel-te-je
I when+INDEF come-FUT-INTR.1SG
'I will come one day.' (Maslova 2003: 168f)

In Example (124a) we have no way of knowing, outside context, whether the event is currently taking place or whether it took place in the past, but in (124b) the event is unambiguously in the future.

Languages may contrast between both before-*Now* and after-*Now*, in which case we also get a specific simultaneous-to-*Now* value. English is an example of a language with a tripartite tense contrast of **past** (before-*Now*) versus **present** (simultaneous-to-*Now*) versus **future** (after-*Now*). I stress again that this refers to grammatical categories, not semantic expressions. A language may well be able to express that an event took place in the past, is taking place now, or will be taking place in the future, without having specifically grammaticalized categories for locating events in time.

TENSE DEPENDENT DEMONSTRATIVES

Chamicuro (Arawakan (Arawakan): Peru) has two clitic particles that function as demonstratives and that contrast for tense, *-na*, used for nonpast contexts and *-ka*, used for past contexts:

p-aškala$^{?}$t-ís-na čamálo
2-kill-2PL-DEF bat
'You (PL) are killing the bat.'

p-aškala$^{?}$t-ís-ka čamálo
2-kill-2PL-DEF bat
'You (PL) killed the bat.' (Parker 1999: 553)

There are also languages without any tense category, where the event is, if needed, located in time through time adverbials or other such strategies. Languages with very little grammatical marking tend to belong to this group. An example of a language without tense is Ju|'hoan.

Ju|'hoan (Khoisan (Northern Khoisan): Angola, Namibia, Botswana)

(125) ha úá Tjùm!kúí
3SG go Tsumkwe
'He went/goes/will go to Tsumkwe.' (Dickens n.y.: 5)

The sentence in (125) cannot be unambiguously translated into English without further context, because English, as opposed to Ju|'hoan, must specify for tense. In other words, without further context than this, all three English translations are equally valid.

For those of us who are used to a tripartite tense system, it might seem highly exotic to have fewer or even none of the values past/present/future. In a pilot survey of 211 languages based on the 222 languages on Östen Dahl's and my tense maps in WALS (Dahl & Velupillai 2011c, Dahl & Velupillai 2011b), I mapped languages as follows: (I) no category tense; (II.a) past tense exists, but not future tense; (II.b) future tense exists but not past tense; (III) both past and future tense exist.[128] While the results in a pilot study are always to be seen as preliminary, they may still serve as a general indication of possible patterns. I found that the majority falls into category (III), both of the category (II) values taken together make up a third of the sample, while the smallest group of languages fall into category (I). Table 8.1 summarizes the findings and Map 8.1 shows the patterns.

Table 8.1 Pilot survey of tense values.

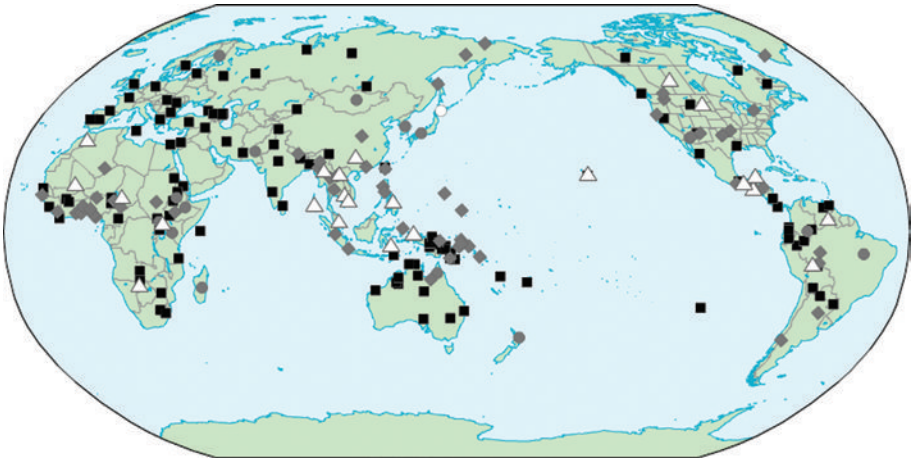
		FEATURE VALUE	#	%
I	△	no tense	24	11.4
II.a	●	past but no future	15	7.1
II.b	◆	future but no past	58	27.6
III	■	both past and future	113	53.8
			210	

As Table 8.1 shows, about one tenth of the languages in the survey lack a specific tense category, while about a third have either a specific past tense or a specific future tense. Of the languages in group (II), it is five times more common to have a future tense but not a specific past tense than the other way about. One might postulate that this could be due to a realis/irrealis dichotomy; it might be more common to need a specific marker for the future since the future has not happened yet and thus belongs to an unreal (irrealis) world, while the past and present both belong to the real world (realis). See 8.5.1 for more on realis and irrealis.

The majority of the languages in the survey have both a past and a future tense. Due to an Indo-European bias, it is likely that the picture is skewed in favour of the group (III) languages. As can be seen on Map 8.1, Australia and the Eurasian landmass are almost entirely dominated by black squares (three tense values), except for the

128. The pilot sample consists of 211 of the 222 languages in our tense maps, 10 of which have been replaced with close relatives.

Eastern edge. The grey symbols (two tense values) form an east-west belt across Africa just under the Sahara, and also a belt stretching north-south across Eastern Asia and the west Pacific from Far Eastern Siberia all the way down northern Australia. The Americas form a more mixed picture, even though the black squares seem to dominate in South America and the grey symbols in North America. As for the two subcategories of value (II), there is no discernible pattern to languages with past but no future tense.



Map 8.1 Pilot survey of tense values. White triangles: no tense (24 languages); grey dots: past/nonpast (15 languages); grey diamonds: future/nonfuture (58 languages); black squares: past/present/future (113 languages). For a full legend, see http://dx.doi.org/10.1075/z.176.additional.

Most languages that lack a specific tense category tend to have a more or less rich aspect system. However, one language in the survey, Maybrat, lacks both tense and aspect categories (Dol 1999).

In Examples (123) and (124) the tenses are marked synthetically with suffixes. Languages may also mark tense analytically with auxiliaries or particles, or non-linearly with stem or tone changes. English is an example of a language that marks future tense analytically, as in *I will cook for you tonight*, where the auxiliary *will* is the future tense marker. Rapanui marks past tense analytically with the particle *i* 'PAST'.

Rapanui (Austronesian (Oceanic): Easter Island, Chile)

(126) a Papi i ma'u i te rama
P.SG Papi PAST take RLT SPE torch
'Papi took the torch.' (Du Feu 1996:156)

An example of tense marked through lexical suppletion (stem change) can be found in English: compare *go* (PRESent tense) versus *went* (PAST tense). Kuche (Niger-Congo (Platoid): Nigeria) is another example of a language where tense is marked through suppletion: compare *āàkɛ̄n* 's/he will run' (FUTure tense) with *áākɛ̄n* 'she ran' (PAST tense), where the difference lies only in tone and stress (Wilson 1996:75).

Languages may also mix strategies; English is an example of such a language, where the past is either synthetically marked (as in *I cook-**ed** for you yesterday*) or marked through lexical suppletion (*go* versus *went*), but where the future tense is marked analytically. It seems more common to have inflectional marking for the past tense than for the future tense: almost exactly half of the languages in our sample, 110 of 222 (49.5%) have inflectional marking for the future tense (Dahl & Velupillai 2011b), while considerably more than half of the languages in Dahl (1985) and Bybee et al. (1994) have inflectional marking for the past tense.

8.2.1 Remoteness

Languages may have more fine-grained grammatical distinctions for time than simply before/simultaneous to/after *Now* in that they grammatically distinguish between degrees of **remoteness** from the *Now*. That is, languages may code an event as having taken place in the near past or remote past, and likewise languages may code for an event going to take place in the immediate or remote future. Of our 222 languages mapped for past tense, 38 (17.1%) have two or three remoteness distinctions, while only two languages (0.9%), Yagua (Peba-Yaguan (Peba-Yaguan): Peru) and Chácobo (Panoan (Panoan): Bolivia), have four or more remoteness distinctions. The biggest group, with 94 (42.3%) languages, have past tense but lack remoteness distinctions, and the second biggest group, with 88 (39.6%) languages, lack a past tense altogether. A very common cut-off point among languages with remoteness in the past is to distinguish between 'today' (**hodiernal** past) or 'not-today' (**hesternal** past). An example of a language with a hodiernal/hesternal system is Supyire, which has three distinct past tense markers: *nî* 'RECent PAST', *ná* 'REMote PAST' and *màha* 'FORmal PAST' (Carlson 1994:329).

Supyire (Niger-Congo (Gur): Mali)

(127) a. mìì nî mu pyi dì yɛ́?
1SG REC.PAST 2SG tell how Q
'What did I tell you (earlier today)?'

b. jò u ná sá lí lwó yɛ?
who 3SG REM.PAST go it take Q
'Who went and took it?' (Carlson 1994:332–3)

The use of the marker *nî* in (127a) means that the event referred to took place before the moment of speech but sometime during the same day. In other words, it would be ungrammatical to combine *nî* with some temporal adverb like *yesterday*. The marker *ná* in (127b) means that the event referred to took place before the current day, which means that it would be ungrammatical to combine it with an adverb like 'today'. The formal past marker *màha* is used to introduce formal narratives, such as folktales and myths.

Languages may code for remoteness in the future too. An example of a language with a bipartite remoteness distinction in the future which is roughly equivalent to a hodiernal/hesternal system is Cupeño, where the immediate future contrasts with the ordinary future.

Cupeño (Uto-Aztecan (Takic): USA)

(128) a. amay=ne aya imi=yaxi-qat mix-an-pi
today=1SG.ERG now 2PL.OBJ=say-IMM.FUT do-*an*[129]-SUBIRR
'Now today I'm going to tell you what to do.'

b. tukumay=ne=pe eme-yka ngiiy
tomorrow=1SG=IRR 2PL-behind go.away.FUT
'Tomorrow I will go after you.' (Hill 2005:128, 130)

The immediate future in (128a) readily combines with such adverbials as 'now', 'today' and 'soon', while the ordinary future in (128b) tends to combine with such adverbials as 'after a while' and 'tomorrow', indicating that it has a higher degree of remoteness. An example of a language with a very complex remoteness system is Yagua, with seven different tenses: irrealis/future, present, proximate-1 (proximate future or immediate past), proximate-2 (proximate future or one day ago past), past-1 (several weeks ago past), past-2 (several months ago past) and past-3 (distant or legendary past) (Payne 1985a: 240).[130] The five remoteness distinctions are shown in Table 8.2.

Some Bantu languages also have highly complex remoteness systems, such as the Kongo (Niger-Congo (Bantoid): DR Congo) dialects Yombe, Western Gogo and Mituku with five remoteness distinctions for the past. To have multiple remoteness in the future is less common. For an in-depth discussion on the tense and aspect systems of Bantu languages, see Nurse (2008).

---

129. *-an* denotes a root augmenting suffix to the verb.

130. Payne (1985a) lists an eighth tense 'narrative present' which essentially is a relative tense where the base form, the present, is used once the time reference has been established at the beginning of the narrative. See further below.

Table 8.2 Remoteness distinctions in the Yagua (Peba-Yaguan (Peba-Yaguan): Peru) tense system (Payne 1985a: 244ff).

TENSE	MEANING	MARKER	EXAMPLE
Proximate-1	immediate future or past	*-jásiy*	rayą́ą́siy ray-jiya-jásiy 1SG-go-PROX1 'I went (this morning).'
Proximate-2	immediate future or one day previous to utterance	*-jay*	rį̨į̨núújeñíí ray-jų̨nnúúy-jay-níí 1SG-see-PROX2-3SG 'I saw him (yesterday).'
Past-1	up to several weeks before utterance	*-siy*	sadííchimyaa sa-díí-siy-maa 3SG-die-PAST1-PF 'He has died (between a week and a month ago).'
Past-2	up to several months before utterance	*-tíy*	sadíítímyaa sa-díí-tíy-maa 3SG-die-PAST2-PF 'He has died (between 1 to 2 months and a year ago).'
Past-3	distant or legendary past	*-jada*	rayúpeeda ray-rupay-jada 1SG-be.born-PAST3 'I was born (a number of years ago).'

8.2.2 Absolute tense

The standard way of defining absolute tense is that it places the **event** or event time (E) either before, after, or simultaneous to the **speech point** (S), i.e. the present moment. We can think of this as E relative S where the event E is placed somewhere on a timeline relative to the speech point S. An absolute **past**, then, would be a grammatical way of expressing that the event is placed before the speech point on a timeline, E before S, as in Figure 8.2[131].

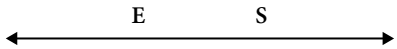


Figure 8.2 Absolute past: the event is placed before the speech point.

131. The figures in this and the subsequent two sections are based on those in Velupillai & Hentschel (2009).

Reading from left to right, the event occurs before the speech point on the timeline. An example of an absolute past would be *I took out the garbage*, where the event (taking out the garbage) took place before the time of my utterance. An absolute **present** places the event at the same time as the speech moment, E simul S, where 'simul' stands for 'simultaneous to', as shown in Figure 8.3.

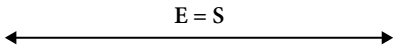


Figure 8.3 Absolute present: the event takes place simultaneously to the speech point.

In Figure 8.3 the event takes place at the same time as the moment of utterance, as indicated by the '=' between E and S. Some languages, English among them, use the present tense to denote generic truths rather than to place an event in the immediate *Now*, for which other constructions are used. In English the progressive (a type of aspect, see below) is used to place an event in the *Now*. Compare *He eats fish* (marked for present tense) versus *He is eating fish* (marked for the present progressive), where the former indicates a general capacity or activity (the person in question can eat fish, i.e. does not shy away from that type of food for any reason) while the latter unambiguously indicates that the act of eating is taking place at the moment of the utterance (the person is at this moment engaged in the activity of eating fish). Many languages, however, only use the regular present tense (without any additional aspect marking) to express the absolute present. An example is Swedish; consider the dialogue below and how it has to be translated into English:

Swedish (Indo-European (Germanic): Sweden)

(129) Q: vad gör du?
what do.PRES 2SG
'What are you doing?'
A: Jag ritar en blomma.
1SG draw.PRES ART flower
'I am drawing a flower.' (source: personal knowledge)

In (129) the question has the verb *göra* 'do' in the present tense because the speaker wants to know what the listener is doing at the time of utterance, S. Likewise the verb *rita* 'draw' in the answer is in the present tense, indicating that the act E is taking place at the time of the utterance. Notice that here the English translations require the progressive aspect (*doing* and *drawing*); it would have been incorrect to translate the question in (129) as **What do you?* or the answer as **I draw a flower*.

The **absolute future** places the event after the speech point, E after S or S before E, both of which are simply alternative ways of expressing the same equation.

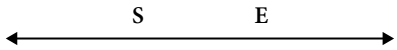


Figure 8.4 Absolute future: the event takes place after the speech point.

An example of an absolute future would be *I will draw a flower for you tomorrow*, where the event (drawing a flower) will take place after the moment of speech.

8.2.3 Relative tense

A relative tense system also places the event on a timeline, but relates it to a given **reference point** (R) instead of to the speech point. The event may thus occur before, simultaneously to or after the reference point. The relative past, **anterior**[132] tense, places the event before the reference point, E before R.

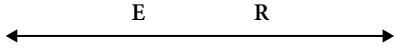


Figure 8.5 Relative past, or anterior: the event is placed before the reference point.

An example of relative past (anterior tense) can be found in Ghanaian Pidgin English, where the location in time of the event is either explicitly established or implied through an already known context.

Ghanaian Pidgin English (Pidgin (English-lexified): Ghana)

(130) ɛnitin hapin fɔ kɔŋkɔmba pipu, dè dè kam kɔmplen
anything happen for Kokomba people 3PB NPU come complain
tu dagɔmba ʧif… bat nau dè wan tu gɛt dɛ̀a
to Dagomba chief but now 3PB INT to get 3PP
on indipɛndɛns
own independence
'(Formerly) the Kokombas referred anything that happened to the Dagomba chief. But now they want to get their independence.' (Huber 1999: 219)

In Example (130) the speaker is first referring to how the state of affairs was before tensions arose between the two ethnic groups (Kokomba and Dagomba), an event in the past that is known to both speaker and hearer; the unmarked verb *hapin* 'happen' thus refers to an event in the past and is located before the subsequently expressed reference point *bat nau* 'but now' on the timeline (which refers to the present state of affairs).

132. Notice that Bybee et al. (1994) use the term 'anterior' to denote what is otherwise called 'perfect'.

It is important to bear in mind that the reference point does not necessarily have to be in the speech moment. In fact, in pure relative tenses, the speech moment is quite irrelevant. We could thus have instances of a relative past occurring prior to the moment of speech, but also posterior to the moment of speech. Compare Figures 8.6 and 8.7:

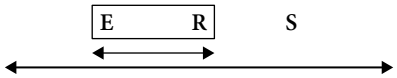


Figure 8.6 The relative past (anterior; E before R) takes place before the speech point.

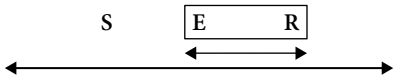


Figure 8.7 The relative past (anterior; E before R) takes place after the speech point.

An example of E before R preceding the moment of speech (Figure 8.6) is, for instance, *I notice that you ate all the chocolate before leaving for school*. Here the reference point R is the act of leaving for school, while E, which comes before R on the timeline, is the act of having eaten all the chocolate. However, both of these acts occurred before the speech point S (*I notice*). An example of E before R coming after the moment of speech (Figure 8.7) would be something like *I am telling you that I will have left when you come home*. Here again the E (my leaving) takes places before R (you coming home), but the whole E before R will only take place after the moment of speech (which is at the time I am telling you about it).

The relative present functions in the same way, namely placing the event simultaneously to the given reference point, E simul R.

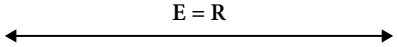


Figure 8.8 Relative present: the event takes place simultaneously to the reference point.

In Figure 8.8 the event takes place at the same time as the reference point. I stress again that the reference point does not have to be equal to the speech point. For instance, E simul R could take place before S. English does not have a grammatical way of marking the relative present, but an example of the principle might be something like *He scratched his head while muttering to himself*, where the act of scratching the head and muttering to himself take place simultaneously, as expressed by the adverbial *while*. Both of these acts, however, take place before the speech point, as

indicated by the past tense *scratched*.[133] And while English does not have a grammatical way of marking it, E simul R may also be placed after the speech point. An example would be something like, *I am telling him the joke and know he will be scratching his head while muttering to himself*, where E simul R (scratching the head at the same time as muttering to himself) will take place after the speech point which is the act of telling the joke.

The relative future, **posterior** tense, locates the event after the given reference point, E after R or R before E.

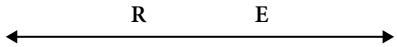


Figure 8.9 The relative future (posterior): the event is placed after the reference point.

English does not have any overt grammatical marking for posterior tense, but a sentence like *He will do it if you ask him* lacks a speech point and so can be said to express relative future where the event (the act of doing something) comes after the reference point (the act of asking). The crucial thing is that the speech point is irrelevant; what is indicated with relative futures is only that E will take place after R. Hausa is an example of a language with a posterior tense, i.e. a relative future tense (Wolff 1993: 431).

8.3 The perfect

The **perfect** is a somewhat more complex category than those described above and has variously been labelled a tense and an aspect category. Recall that the tense categories above all have the basic principle of locating an event before, simultaneously to, or after a given location on the timeline (either the speech point or a reference point). A very essential and crucial property of the perfect, however, is that while it places an event prior to a given location on the timeline, the event is *still relevant* at that location. In other words, the perfect spans two separate locations on the timeline, since, even though the event itself took place before the speech or reference point, it is still valid at S or R, that is, it is still having an effect on or is somehow relevant to the given reference point. Consider Figure 8.10:

---

133. The **historic present** could also be argued to be a relative tense: something like *Marie Antoinette kneels on the guillotine and the crowd cheers* would be E=R but the whole episode is clearly located before the moment of speech.

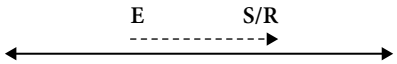


Figure 8.10 Perfect tense: the event takes place before the speech/reference point, but is *still relevant* at that point on the timeline.

In Figure 8.10 the event is located before the speech/reference point on the timeline, but the dotted arrow indicates that E is still relevant at the speech/reference point. Consider Pooh's reply to Owl's question of how he is doing:

(131) "Terrible and Sad," said Pooh, "because Eeyore, who is a friend of mine, has lost his tail. […]" (Milne 1974: 46)

In Example (131) the perfect *has lost* indicates that the event took place before the time of Pooh's recounting it – the disappearance of the tail happened at some point in the past in relation to the *Now* on Pooh's timeline – but it is still relevant when Pooh mentions it; the tail is still lost.

The entire equation of an event that occurs before a given point but that is still relevant at that point can be placed before or after the speech point on a timeline to form the **past perfect** (**pluperfect**) or the **future perfect**. These can in effect be said to span over three separate locations on the timeline: the E, which takes place before R (but is still relevant at the point of R), all of which is placed in relation to S.[134] In the case of the pluperfect, the perfect is placed prior to the speech point, as in Figure 8.11:

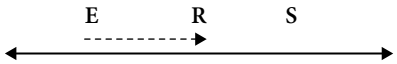


Figure 8.11 Pluperfect tense: the event takes place before the reference point and is still relevant at that point on the timeline, all of which is located before the speech point.

An example of the constellation shown in Figure 8.11 would be something like *I had already sealed the envelope when I saw the check lying on the table so now I have to open it again*. Here the event at the speech point is in the *Now*, thus in the present tense (*have*), while the event situated prior to the *Now*, namely R, is in the past (*saw*). The event that stands in relation to R and is also relevant at R, namely E, is in the pluperfect (*had sealed*), since the envelope was still sealed at the time I saw the check.

The future perfect functions in the same way, except that the speech point moves to before E, as in Figure 8.12.

---

134. It could thus be argued that pluperfect and future perfect are absolute tenses, since they are by necessity related to the speech point.

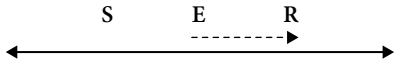


Figure 8.12 Future perfect tense: the event takes place before the reference point and is still relevant at that point on the timeline, all of which is located after the speech point.

An example of future perfect would be something like *He says that he will have finished before we arrive*. Here *says* is in the present tense indicating the speech point. At a later stage on the timeline, after the speech point, the E (*he will finish*) will be something of the past at the time of R (*we arrive*) but will still be relevant at R (he will still be finished).

The above discussed use of the perfect may also be termed **resultative**. If a perfect is used to describe an event that took place repeatedly or over a span of time, it may be labelled as **experiential** (Dahl & Velupillai 2011d). An example would be something like *It has been known to happen before*, where the speaker indicates that whatever happened now is not unique, because it has happened repeatedly in the past, and the experience of those occasions still holds. In our survey we noted languages as having the perfect tense if the language included both the resultative and experiential use. This means that those languages where the construction is only used as a resultative or only as an experiential were listed as not having the perfect. Slightly more than half of the languages in our sample, 114 of 222 (51.4%) lack the perfect. While these languages are spread over the world, they cluster somewhat in Australia as well as North and South America.

An interesting feature of the perfect is where it derives from. While those of us who are used to mainly Western European languages might assume that the default is to derive the perfect from the possessive construction, as it is in English (*have* VERB), this in fact a very rare phenomenon found in only seven of the 118 languages that have perfect (5.9% of the languages with perfect or 3.2% of the languages in the entire sample). It is, in fact, in our sample found only in Indo-European languages located in Europe on the map: English, German, Icelandic, Swedish, French, Spanish and Modern Greek. Three times as many languages, 21 (17.8% of the languages that have perfect or 9.5% of the whole sample), derive their perfects from expressions meaning 'finish' or 'already'. The remaining 80 languages with perfects have various other sources for the construction. For more on the origins of grammatical markers (grammaticalization), see 13.1.

Another interesting feature of the perfect is that it may develop into a simple past tense (E before S) or a perfective aspect (see below); this has happened or is currently happening in central Western Europe, in an area stretching from Germany down to Corsica and from the Atlantic coast of France and Belgium to the eastern border of Germany and Austria (Dahl & Velupillai 2011d).

8.4 Aspect

While tense is the linguistic tool for placing an event on a timeline, **aspect** can be said to be, very simplified, the tool for defining the perspective taken on an event. In other words, aspect can be thought of as the device used to grammatically express different views of events in relation to their respective start and end points. For thorough discussions on aspect, see, for example, Comrie (1976) and Dahl (1985), which are classics by now. For a more theoretical – but very elegant – framework for aspect, see Smith (1997) and Johanson (2000).

It is important not to confuse tense with aspect, even if they are interrelated in many ways. Irrespective of what perspective we take on an event it can be placed on a timeline relative to a given reference point: *He coughed* (once) versus *He was coughing* (repeatedly or over a period of time) are two different perspectives taken on the same event, both of which are located at a particular point on the timeline (E before S/R). The location on the timeline can change, however, even if the respective perspectives do not: the sentences *He will cough* (once) versus *He will be coughing* (repeatedly or over a period of time) keep the perspectives intact, but move the event to a different location on the timeline (E after S/R).

8.4.1 Aspect versus *Aktionsart*

While aspect deals with the perspective taken on an event, ***Aktionsart*** (also called **actionality**, **lexical aspect** or sometimes **derivational aspect**) specifies the inner structure of the event (see, for example, Johanson 2000 for a very thorough discussion on the difference between aspect and *Aktionsart* or actionality). Aspect is a grammatical category while *Aktionsart* is a lexical, semantic, specification. Consider the following verbs: *seethe*, *dwell*, *cough*, *freeze*, *sing*, *build*. They can all be placed at different locations on the timeline relative to a reference point: *seethed*/*will seethe*, *dwelled*/*will dwell*, *coughed*/*will cough*, *froze*/*will freeze*, *sang*/*will sing*, *built*/*will build*. Yet they are inherently different in many ways. This is because they differ in their *Aktionsart* or actionality. There is something inherently dynamic about *seethe* which is not there in a verb like *dwell*. There is something inherently punctual about *cough* which is not there in *freeze*. And there is an inherent end point in *build* which is not there in *sing*.

Verbs may be **dynamic** or **stative** (non-dynamic), which means that their inherent semantics encode whether their inner structure involve any form of change or not. With dynamic verbs there is an inherent element of change: if something seethes or runs or falls the event itself encodes some kind of action. With stative verbs there is no element of change, merely a constant state: if someone knows (something) or dwells (somewhere) or something contains (something) there is no action involved and there is no inherent meaning of internal evolution.

Verbs may also be **punctual** or **durative**, which means that their inherent semantics encode whether their inner structure allow for a duration in time or not. With punctual verbs there is no real internal structure to the event: to cough or to sneeze or to flash are all events that last only an instant and have no particular inner structure consisting of different phases of that event. Durative verbs, on the other hand, do contain this inner structure that is made up of a string of phases: if something freezes or burns or blows, there is an element of time inherent in the verb. The various phases that are inherent in the verb freeze, from the stage where the element is completely unfrozen, to the stage where it is completely frozen, involve a gradual change which in itself implies a certain amount of duration. Notice that dynamicity and punctuality are not necessarily mutually exclusive: an example of a dynamic punctual verb would be *cough*, while *seethe* is a dynamic durative verb. Stative verbs tend to be durative.

Finally, verbs may be **telic** or **atelic**, which means that their inherent semantics imply an inherent end point (telic) or no inherent end point (atelic). Telic verbs inherently contain an element of the action or event coming to an end: *to build*, *to make*, *to bake* are all verbs that imply that even if the action has a certain amount of duration, eventually it will end, since eventually we will have finished building the house or finished making the pot or finished baking the cake, unless we break off in the middle. If we break off in the middle, however, we cannot say that we have built or made or baked something. Atelic verbs do not contain this inherent end point: even if such verbs as *sing* or *play* or *dance* involve a certain duration in time, there is no implicit finishing point. We may stop singing or dancing or playing, but that doesn't leave any unfinished business as it would if we broke off in the middle of building or baking or making something.

8.4.2 Perfective versus imperfective

As mentioned above, the traditional definition of aspect is that it deals with the internal structure of an event. But as we have seen, this is rather a definition of *Aktionsart* or actionality. Aspect, on the other hand, is a grammatical category denoting the perspective taken on an event. That is, it deals with "morphological devices expressing different views of linguistically represented events, envisaging them in various ways relative to their limits, and signalling how they come into view at aspectual viewpoints" (Johanson 2001:7). This section will give a very simplified and brief definition of the two major aspect categories, perfective and imperfective, as well as some common subcategories. In essence, this is a simplified version of Johanson's (2000) framework.

An event may have a beginning, a course and an end. The event in question may be viewed from within, during its course, or from outside its course, as a demarcated or bounded whole. The basic notion of **perfective** aspect is that the perspective is on

the whole, bounded or 'captured' event, while the basic notion of the **imperfective** aspect is that the perspective is within the event, during its course. Figure 8.13 is an attempt to illustrate the basic difference between the two:

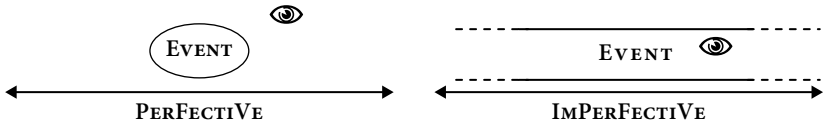


Figure 8.13 The basic difference between perfective and imperfective aspect is that with perfective aspect the perspective (illustrated with the eye symbol) is on the event as a bounded whole, while with imperfective aspect the perspective is within the event, during its course.

It is very important here to not confuse the perfect tense with the perfective aspect despite the similarity of the two terms. The former locates an event on the timeline, while the latter specifies how the event is viewed (namely as a bounded whole). Viewing an event as a bounded whole does not necessarily tell us much about the internal structure of that event. It may be a punctual event with no internal structure at all, or it may be a durative dynamic event with an internal structure. The crucial thing is that the event is viewed in its totality, as a bounded unit. Viewing an event from within also tells us rather little about the internal structure of that event. Even if it is an event of a very limited durativity, it can, theoretically, be viewed from within. Most commonly, however, punctual events combined with an imperfective aspect tend to denote repetitions of the event.

An example of a language with a bipartite perfective/imperfective system is Rendille, where the ending indicates aspect:

Rendille (Afro-Asiatic (Eastern Cushitic): Kenya)

(132) a. khadaabbe chiirte
letter.PL write.PFV
'He wrote letters.'

b. khadaabbe chiirta
letter.PL write.IPFV
'He writes/is writing/wrote/was writing/will write letters.'

(Dahl & Velupillai 2011a)

In (132a) the ending *-e* signals perfective aspect and specifies that the event is a bounded whole. In (132b) the ending *-a* signals imperfective aspect and indicates that event is viewed from within, during its course. There is neither a beginning point nor an end point indicated in (132b). In fact, there is no specification whatsoever, except that the event is (or was or will be) taking place.

The perfective aspect readily combines with the past tense, since for an event to be viewed in its totality it typically has to have come to an end. Notice, however, that strictly speaking that is aspect *in combination* with tense, in the sense that the event and the perspective taken on it are placed prior to a given reference point on the timeline. We may thus also have the logical combination of an event in the imperfective aspect placed prior to a given reference point on the timeline. An example of the contrasts perfective versus imperfective in combination with the past tense can be found in literary French, where an event located E before S on the timeline may be either in the *passé simple* (past definite) or in the *imparfait* (imperfect tense or past indefinite):

French (Indo-European (Romance): France)

(133) a. Marie mangea lentement
PN eat.PASTD slowly
'Marie ate slowly.'

b. Marie mangeait lentement
PN eat.PASTI slowly
'Marie ate slowly.' (source: personal knowledge)

The difference between the sentences in (133a and b) is one of aspect: the events are both located in the past but in (133a) the event is a bounded whole and urges the narrative forward – we want to know what happened next – whereas in (133b) the event is not bounded and serves as a background for other things that happened while Marie was eating. This is a typical difference between perfective and imperfective aspects: the perfective tends to drive a narrative forward, viewing each event in its totality, while the imperfective tends to serve as a background against which other events take place. It is not possible to completely capture in English the aspectual difference between the two sentences in (133); the nearest translational equivalents would be something like 'Marie ate slowly' (then something else happened) for (133a) and 'Marie was eating slowly' (while something else occurred) for (133b).

While it is less common to combine the perfective/imperfective opposition with the future tense than it is to combine it with the past tense, there are languages with this opposition. Modern Greek, for example, has two different forms for the future, the future aorist (future tense with perfective aspect) and the 'future paratatikos' or continuous future (future tense with imperfective aspect). The difference between *grapso* (write.1SG.FUT:AOR) 'I will write' and *grapho* (write.1SG.FUT:PAR) 'I will write' is that the former event is viewed as a bounded whole placed in the future on a timeline, while the latter, also placed in the future on a timeline, is viewed from within. The formed denotes that there will be an end to the event after which the next event will take place, while the latter has no such connotation and merely indicates that the writing event will be taking place in the future, and might be better translated as 'I will be writing (when X, Y or Z will take place)'.

It is quite common in the world to have grammatical marking of perfective versus imperfective aspect. In our sample of 222 languages 101 (45.5%) have grammatical marking of aspect while 121 (54.5%) do not (Dahl & Velupillai 2011a). French is an example of the former and English an example of the latter. Geographically the picture is rather mixed, but Northern Europe and South-East Asia are areas where languages consistently lack grammatical aspect.

The **progressive** aspect can be thought of as a subcategory of the imperfective. It specifically denotes that the event is ongoing. This is the only inflectionally marked aspect category that English has, the *-ing* form. This may combine with tense to form either the past, present or future progressive by combining with a tense inflected form of BE (as in *I was writing*, *I am writing*, *I will be writing*). The *Aktionsart* or actionality of the verb is largely irrelevant, although progressive aspect with a stative verb tends to get a special reading of confinedness. So while someone can *be running* (progressive aspect, dynamic verb), *be blowing* (progressive aspect, durative verb), *be building* (progressive aspect, telic verb) and *be singing* (progressive aspect, atelic verb), to say that someone *is knowing*/*containing* something would only be possible in highly specific contexts. The progressive aspect with a punctual *Aktionsart* tends to give an iterative meaning in English: *The lamp is flashing* implies that the punctual act of the flash is repeated several times.

The **habitual** denotes that an event takes place regularly or is true for an extended period. As such it can also be thought of as a subcategory of the imperfective. The habitual should not be confused with iterativity, which specifically denotes repetition on a single occasion. The habitual merely states that an event is true over an extended period of time. English has a tense restricted habitual for the past only: the construction *used to* VERB, which indicates that the event was true for an extended period of time to the degree that it becomes characteristic of this period, as in *I used to live in Berlin* or *I used to play the piano every afternoon* or *I used to cough in the nights when I was a child*. An example of a language where the habitual does not have tense restrictions is Nalik.

Nalik (Austronesian (Oceanic): Papua New Guinea)

(134) gu runa va-nam-doxo yang ni
2NPL HAB CAU-stomach-good FOC 1SG
'You (always) make me so happy.' (Volker 1994: 80)

In (134) there is no tense specified but simply the fact that the event happens often over a long period of time.[135]

135. This would be termed habitual-generic (HABG) in Dahl (1985), where a discussion on the difference between habitual and generic is given.

The **iterative** denotes repetition (iteration) of an event on a single occasion, indicating that the event took place as a series of bounded wholes. As such it might conceptually be considered a subcategory of perfective aspect, in the sense that (*i*) it refers to a series of repeated bounded events and (*ii*) this series of repeated bounded events occur on one single occasion. In other words, it is a bounded event consisting of several individually bounded events. However, it tends to be expressed through imperfective means. English does not have a specific form for iterative aspect, and instead expresses it with such expressions as *to* VERB *and* VERB, or *to* VERB *again and again* as in *He blew and blew* or *He dialled the number again and again (but they never answered)*. The iterative is often marked through reduplication, as in Kayardild, where verbal reduplication indicates multiple repetition of an event:

Kayardild (Australian (Tangkic): Australia)

(135) dara~dara-tha raa-ja warirr
break-RED-ACT spear-ACT nothing
'(They) speared (him) but (their spears) broke and broke again, nothing (happened).' (lit. 'Speared, broke and broke, nothing.')

waldarra jabi~jabi-j kurumbu bula-a-nangku
moon.NOM shudder-RED-ACT barbed.spear.NOM pull-MOD-NEGPOT
'Moon shuddered and shuddered but the spear could not be pulled out.'
(Evans 1995: 290)

In (135) the reduplicated verbs *dara-* 'break' and *jabi-* 'shudder' denote that the event is repeated on a single occasion. The repeated breakings and shudderings are not spread out over an extended period of time, but happen several times on one occasion.

The **completive** denotes the completion of an event and as such could be argued to be a kind of subcategory of the perfective aspect. English does not have a specific completive marker, and instead has to use such expressions as *to finish* VERB, as in *He finished analysing his data*. Completive aspect may often be translated into the English perfect (*X has happened*). Engenni is an example of a language with completive, as is Hawai'i Creole English.

Engenni (Niger-Congo (Edoid): Nigeria)

(136) á gbè ạ̀dhè bhi nì o
one let.go.home day be.black COMPL in.fact
'Let's go home! It has got dark, you know.' (Thomas 1978: 73)

Hawai'i Creole English (Creole: Hawai'i, USA)

(137) da wahinɛ… ða wan dæd ʤɛs paʊ hanaʊ
DEF woman DEF INDEF REL just COMPL give.birth
'The woman (…) the one that just gave birth.' (lit. '…the one that just finished give.birth.') (Velupillai 2003: 98)

The completive markers in the examples above denote that the event has reached completion (*nì* in Engenni and *pau* in Hawai'i Creole English). Notice that this is not possible to capture exactly in English; the fact that Example (136) translates into an English perfect and Example (137) into an English past is only because those are the closest equivalent meanings that allow for a fluent English sentence and not because these markers denote the perfect or past tenses.

8.5 Mood and modality

Mode is, very simplified, the category that codes a speaker's attitude toward a situation or statement. This includes the speaker's belief in the reality of the event, the likelihood that the event will occur, or the quality of information that the speaker has about the event. Once again, it is important to keep in mind that we are dealing with grammatical categories. Many languages will have ways of expressing attitudes and beliefs towards an event, but they may or may not have grammatical categories specifically used to code these attitudes. The terms 'mood' and 'modality' are often used interchangeably. When they are distinguished, **mood** tends to denote a higher level distinction for the whole clause of realis (asserting that a proposition holds true) versus irrealis (making no assertion with respect to the truth of the proposition) while **modality** denotes semantic labels of attitudes towards events. The term 'mode' is sometimes used to cover both mood and modality. For a very accessible and thorough discussion on mood and modality, see Palmer (2001), which this section relies heavily on.

8.5.1 Realis and irrealis

Realis is typically used when the speaker is very sure that the event has happened or that the state of affairs holds true. **Irrealis**, on the other hand, carries no such assertions. It is basically a distinction between "actual and non-actual events" (Chung & Timberlake 1985: 241), or between asserted and non-asserted propositions. A sentence like *I rang the doorbell* would be an example of something in the realis; I am asserting as a true and indisputable fact in the real world that I rang the doorbell (or at the very least, I am utterly convinced about the veracity of that fact). A sentence like *If I ring the doorbell*, however, would be an example of irrealis. The event has not happened and might never happen in the real world, hence an assertion of absolute certainty about the fact of the event would be inappropriate. Notice here that irrealis does not necessarily state that the event will not take place or is not true, it simply "makes no claims with respect to the actuality of the event or situation described" (Payne 1997: 244). Consider the following examples from Tugun:

Tugun (Austronesian (Central Malayo-Polynesian): Indonesia)

(138) a. ra-tunu ika
3PL.R-cook fish
'They cooked fish.'

b. lalaik Ø-gisan hala? lalaik Ø-gisan raha
man 3SG.R-make what man 3SG.R-make house
'What are you doing? I'm working on my house.'

c. au mu-la naha
1SG 1SG.IRR-go first
'I'll go now.'

d. hira marr-ala la hira
3PL 3PL.IRR-take to 3PL
'They wanted to take (them) for themselves.'

e. mu-osi oni hala le mu-seli
1SG.IRR-do like what so 1SG.IRR-pry.up
'What should I do in order to pry (it) up?'

f. om-oci ni heri mu-ta-mate
2SG.IRR-use 3SG DEM 1SG.IRR-NEG-die
'If you use this, I won't die.' (Hinton 1991: 99)

The unifying element of Examples (138a and b) is that it is an undisputable fact that they happened or are happening, indicated by the portmanteau markers Ø (3SG.Realis) and *ra-* (3PL.Realis) while the unifying element in Examples (138c–f) is that an assertion of undisputable reality of the events or situations is not possible, and thus the verbs have to be coded in the irrealis mode, indicated by the portmanteau markers *mu-* (1SG.IRRealis), *marr-* (3PL.IRRealis) and *om-* (2SG.IRRealis). Notice, however, that this does not mean that the speaker is claiming that the events or situations will never happen, but simply that it is not possible to absolutely assert that they have taken place (or are taking place) in the real world. In Examples (138c and f), for instance, it seems safe to assume that the speaker is quite convinced that the respective events will take place. Likewise in Example (138e) it is reasonable to assume that the speaker is convinced that the event of prying open will take place once the relevant advice of how to do it has been given. Example (138d) is a counterfactual: the taking (of something for themselves) never took place in the real world.

The realis/irrealis opposition is essentially the same as the indicative/subjunctive opposition found in many Indo-European languages, such as in the Spanish example below.

Spanish (Indo-European (Romance): Spain)

(139) si no hubiera sido por Anita, mi reloj sería perdido
if NEG have.SUBJ been for PN my watch be.SUBJ lost
'If it had not been for Anita, my watch would be lost.' (Payne 1997: 245)

GRAMMATICALIZED FUTILITY

In Mapudungun (Araucanian (Araucanian): Chile) the suffix *-fu* denotes 'ruptured implicature' (RI), meaning that an expected result did not occur, so the event or action was to no avail.

amu-fu-n tañi weŋüy mew, welu pe-la-fi-ñ
go-RI-1SG 1SG.POSS friend PPOS but find-NEG-DIR-1SG
'I went to my friend's (house), but did not find him/her.'

kiñe küyeŋ dewma mawüŋ-fu-i
one month already rain-RI-IND
'It rained a month ago (but to no avail [the grass did not get green].' (Zúñiga 2000: 45)

For a discussion on the difference in function between the irrealis and subjunctive modes, see Palmer (2001).

Realis mood interacts readily with such tenses as the past, present and nonfuture, i.e. tenses that place the events on such a location on the timeline that an assertion about their validity is possible. If something has already taken place, or is right now taking place, it is possible to assert that the proposition holds true, which makes the realis mood appropriate to use. Irrealis mood, on the other hand, tends to interact readily with such tenses as the future and nonpast, i.e. tenses that place the events on such a location on the timeline that an assertion about their validity is not possible. I stress again that this does not necessarily mean that the events are never going to happen, merely that it is not possible to make any claims about their validity. If something is predicted to happen in the future, the speaker may be utterly convinced that it *will* happen, but will still not be able to assert that it *has* happened. Consider the examples from Anjam below:

Anjam (Trans-New Guinea (Madang): Papua New Guinea)

(140) a. e tabir yans-eqn-a-m Rut alaŋ-oqn-e-j
1SG dishes wash-SIM.R-REM.PAST-1SG.R.DS PN play-CONTR-REM.PAST-3SG
'While I washed the dishes Ruth played.'

b. e ino bem qoit-et-i-t ni uy-e
1SG your bread bake-BEN-FUT-1SG.IR.DS 2SG eat-IMP
'When I bake your bread, you eat.' (Roberts 1990: 382)

In Example (140a) the simultaneous realis marker *-eqn-* is used with the remote past marker *-a-* and the suffix *-m*, the first person singular, **realis**, different subject marker; because the event has already taken place, it is quite possible to assert its veracity. In (140b) the future marker *-i-* is used with the suffix *-t*, the first person singular, **irrealis**, different subject marker; because the event hasn't taken place yet, it is not possible to absolutely assert its factuality, even if it is highly likely to happen.

8.5.2 Propositional modalities

Propositional modalities have to do with speakers' attitudes towards the truth value of the information given in the proposition. There are two basic subcategories of propositional modality, evidentials and epistemic modalities. **Evidentials** code the type of evidence a speaker has for a given proposition, while **epistemic** modalities code qualitative judgements about the information in a given proposition.

8.5.2.1 *Evidentials*

The type of evidence had for a proposition can be either **sensory** (or **direct**), where a speaker is indicating that the evidence had for a proposition is through evidence of the senses (usually auditory or visual), or the evidence can be **reported** (or **indirect**), where a speaker is not a witness to the event but has the evidence for the proposition through hearsay or similar sources. For a thorough discussion on evidentials, see Aikhenvald (2004).

Sensory or direct evidentials are usually based on visual evidence, i.e. that the speaker saw it with his/her own eyes. Reported or indirect evidentials can be either **inferential**, in which case the speaker infers the truth of a proposition based on physical evidence, or they can be **quotative** (also referred to as 'hearsay', 'reportatives', or 'second-hand evidentials'), in which case the speaker has been told about an event. Qiang is an example of a language that has all three kinds.

Qiang (Sino-Tibetan (Quiangic): China)

(141) a. the: jimi de-se-ji-w-a
3SG fertilizer OR-spread-CSM-VIS-1SG
'She spread the fertilizer (I saw her do it).'

b. panə-le ɦa-χ̊ə-k-ən
thing-DEF OR-broken-INFR-2SG
'It seems you broke the thing.' (Inference from seeing the broken pieces in the person's hand.)

c. the: ʑdʑyta: ɦa-qə-i
3SG Chengdu.LOC OR-go-HS
'He went to Chengdu (I heard).' (LaPolla 2003: 66, 64, 70)

In Example (141a) the suffix *-w* indicates that the speaker saw the event take place, while in (141b) the suffix *-k* indicates that speaker only has indirect evidence for the truth of the proposition, namely the physical evidence that presumably came about as a result of the event. In (141c) the suffix *-i* indicates that the speaker is reporting what s/he has been told. Foe is an example of a language with a very complex evidentiality system:

Foe (Trans-New Guinea (Kutuban): Papua New Guinea)

(142) a. na mini wa-bugege
I today come-PRES.PARTICIPATING.EV
'I am coming today.' (I am participating in the action or making a statement of a generally known fact.)

b. aiya baye wa-boba'ae
airplane come-VIS.EV
'An airplane is coming.' (I can see it.)

c. aiya baye wa-bida'ae
airplane come-NONVIS.EV
'An airplane is coming.' (I can perceive it non-visually by hearing, smell, feeling or understanding.)

d. Kabe Irabo wa-ada'ae
Mr. PN come-DEDUCTIVE.EV
'Mr. Irabo is coming.' (Inference based on something for which I have evidence perceived with my senses – e.g. I can hear Mr. Irabo and can recognize his voice.)

e. Kabu Maduane minage wa-bubege
Mr. PN still come-PREVIOUS.EVIDENCE.EV
'Mr. Maduane is still coming.' (I have seen the evidence but cannot see it at the moment of speech – e.g. we both left together but I was faster than Mr. Maduane, but I know he's still on his way and will be coming.)

(Aikhenvald 2004: 62 citing Rule 1977: 71)

An example of a language with a very rare system indeed is Matsés, where tense and evidentiality interact in so far as that speakers must obligatorily specify two tenses, the first one indicating when an inferred event happened, and the second one indicating when the evidence for the inference was encountered, termed 'double tense' by Fleck (2007).

Matsés (Panoan (Panoan): Brazil, Peru)

(143) a. mayu-n bëste-wa-ak-onda-ṣh
non.Matsés.Indian-ERG hut-make-REC.PAST.INFER-DIST.PAST.EXP-3
'Non-Matsés Indians (had) made a hut.'

b. mayu-n bëste-wa-nëdak-o-ṣh
non.Matsés.Indian-ERG hut-make-DIST.PAST.INFER-REC.PAST.EXP-3
'Non-Matsés Indians (had) made a hut.' (Fleck 2007: 589–90)

The difference between (143a and b) is that the former specifies that the speaker discovered the hut a long time ago (*-onda-* 'distant past experiential'), but at the time of the discovery the hut had only recently been made (*-ak-* 'recent past inference'), whereas the latter specifies that the speaker recently discovered the hut (*-o-* 'recent

past experiential') which at the time of the discovery was old, i.e. had been made a long time ago (*-nëdak-* 'distant past inference'). There are three remoteness distinctions in Matsés – recent past (immediate past to about one month ago), distant past (about one month ago to about 50 years ago) and remote past (more than 50 years ago) – and three kinds of evidentials (experiential, inference and conjecture). There are thus nine logical combinations for the markers, although there are two 'remote past inferential' markers but no 'remote past conjecture' marker (Fleck 2007).

In de Haan's (2011) sample it is more common for languages to have evidentials than not. Of his 418 languages, 237 (56.7%) have evidentials while 181 (43.3%) lack them. English belongs to the latter group. Of the languages that have them, it is far more common to only have indirect evidentials (found in 166 or 70% of the languages with evidentials) than to have both direct and indirect evidentials (found in 71 or 30% of the languages with evidentials). While evidentials are spread over the world, they are extremely abundant in North and South America and quite rare in Africa.

8.5.2.2 *Epistemics*

Epistemic modalities code the speaker's qualitative judgement of the proposition. **Speculative** judgements indicate that the speaker is uncertain about the factual status of the proposition, as in *Peter **may** be home already*, where *may* serves to indicate that the speaker is not certain about whether the proposition holds true. **Deductive** judgements indicate that the speaker is inferring something on the basis of external evidence, as in *Peter **must** be home already (since the lights are on)*. Deductive judgements tend to imply that the speaker is making a firm judgement; s/he is quite convinced of the truth of the proposition, while a speculative judgement implies a lesser degree of conviction. An **assumptive** judgement denotes that the speaker is inferring something on the basis of what is generally known, as in *Peter'**ll** be home already (because he's always home before six o'clock)*. Here the speaker is making an assumption based on what is generally known about Peter's habits. English does not have a specific marker for assumptive, but expresses the equivalent with the future tense marker (*will*) or with the present tense. Wintu is an example of a language with a specific marker that could be glossed as 'assumptive':

Wintu (Penutian (Wintuan): USA)

(144) pi kupa-ʔel
he chop-EXPECT[136]
'He is chopping wood.' (Pitkin 1985:135)

---

136. Pitkin terms *-ʔel* an "experiential evidential suffix" (1985:135) while Schlichter (1986:51) terms it 'EXPECTational'; I am thus using Schlichter's gloss. Both Pitkin and Schlichter analyse the suffix as an evidential marker.

The suffix -*ʔel* in Example (144) indicates that the speaker expects, or assumes, that the proposition holds true on the basis of general knowledge. For example, it would be justified to state that *He is chopping wood*, using the expectational (or assumptive) suffix -*ʔel* if "he has a job cutting wood, he usually goes every day between 8 and 5, it is 3 o'clock and yesterday at 3 o'clock he was chopping wood" (Pitkin 1985: 135) or something similar. Notice that this can be analysed as an evidential (and in fact is in the sources), showing that there are no sharp boundaries between evidentials and epistemic modalities.

8.5.3 Event modalities

Event modalities have to do with potential action in various ways. The potential events or actions have not been realized yet, but are possibilities or probabilities. There are two subcategories of event modality, deontic and dynamic modalities. With **deontic** modalities external factors (such as obligation) are those initiating or conditioning the action, while with **dynamic** modalities internal factors (such as ability) are those initiating or conditioning the action.[137]

8.5.3.1 *Deontics*

Deontic modalities have two further subgroups, directives and commissives. **Directives** are used when the speaker tries to initiate action, such as stating an obligation, giving permission or giving an order (an imperative)[138]. Commissives, on the other hand, are used when the speaker certifies that an action will take place. An example of an imperative would be something like *Give me that!* or *Walk!*, where the speaker is giving the addressee a direct order. For a very detailed study on the forms and functions of imperatives and commands, see Aikhenvald (2010).

English does not have a special morphological form for the imperative,[139] but in fact it is much more common to have one: 425 of 547 languages (or 77.7%) in van der Auwera & Lejeune's (2011a) sample have a special morphological form for the imperative while 122 (22.3%) do not. Most commonly, languages have different morphological forms depending on the number of the addressee(s), for example, one for the second person singular (one addressee) and one for the second person plural (more than one addressee). This is found in 292 languages in the sample (68.7% of the languages that

---

137. The term **situational** modalities covers deontic obligations and permissions and well as dynamic abilities, cf., for example van der Auwera & Ammann (2011b).

138. Termed 'speaker oriented' modalities in Bybee et al. (1994).

139. Prosody, however, clearly marks off an imperative from a present tense form.

have special morphological forms for the imperative or 53.4% of the entire sample). An example of such a language is Comanche, where the imperative has three different morphological forms.

Comanche (Uto-Aztecan (Numic): USA)

(145) a. yu-kat*ɨ*
quiet-sit(SG.SUBJ)
'Sit down and be quiet!' (said to one addressee)

b. yu-y*ɨ*kwi-p*ɨ*kw*ɨ*h
quiet-sit(PL.SUBJ)-DU.IMP
'Sit down and be quiet!' (said to two addressees)

c. yu-y*ɨ*kwi-ka
quiet-sit(PL.SUBJ)-PL.IMP
'Sit down and be quiet!' (said to more than two addressees)

(Charney 1993: 217)

In (145) the imperative differs in two respects according to the number of addressees: the stem varies according to whether the subject is singular or not (*kat*ɨ 'sit' is used for singular subjects while *y*ɨ*kwi* 'sit' is used for non-singular subjects); also, an imperative suffix indicating the number of the subjects (*-p*ɨ*kw*ɨ*h* for the dual and *-ka* for the plural) is required for the non-singular imperatives.

In 42 languages (9.9% of the languages with a special morphological form for the imperative and 7.7% of the entire sample) there is a special morphological imperative form for the second person singular, but not for any other number. An example of such a language is Italian, where the imperative takes a special morphological form for one addressee, but where it looks like the present tense form if there are two or more addressees:

Italian (Indo-European (Romance): Italy)

(146) a. parl-a
talk-IMP.2SG
'Talk!'

b. parl-ate
talk-PRES.IND.2PL
'Talk!'/'You (PL) talk.' (source: personal knowledge)

A very rare strategy indeed, found in only two languages (0.5% of the languages with a special morphological form for the imperative and 0.4% of the whole sample), namely Latvian and Apurinã (Arawakan (Arawakan): Brazil), is to have a special morphological form for the second person plural, but not for the second person singular.

Latvian (Indo-European (Baltic): Latvia)

(147) a. dzer
drink.PRES.IND.2SG
'Drink!'
b. dzer-ie-t
drink-IMP-2PL
'Drink!' (Holst 2001:179)

Another strategy, found in 89 languages (20.9% of the languages with a special morphological form for the imperative and 16.3% of the entire sample) is to have a special morphological form for the imperative that is number neutral, as is the case in Swedish:

Swedish (Indo-European (Germanic): Sweden)

(148) a. ät!
eat.IMP(SG/PL)
'Eat!'
b. du/ni ät-er
2SG/2PL eat-PRES(SG/PL)
'You (SG/PL) are eating.' (source: personal knowledge)

The bare form in (148a), *ät-Ø* 'eat', serves as an imperative irrespective of the number of addressees, while the form in (148b), *äter* 'eat', carries the present tense suffix *-er*.

The two other directives involve stating obligations or permissions. An example of a deontic **obligative** would be something like *You **must** send the letter*, where *must* denotes that the speaker is trying to initiate action with the addressee by uttering an obligation. An example of a **permissive** would be something like *You may (/can) go to the party*, where *may* (or *can*) indicates that the speaker is giving the addressee permission for an action. Notice that in English the same markers are used to code epistemic speculative and deductive judgements as to code deontic obligations and permissions. Thus *must* is used both for obligations (*You **must** go now*) and for deductive judgements (*He **must** be home already*) and *may* is used both for permissions (*You **may** go now*) and for speculative judgements (*He **may** leave tomorrow*). To have such an overlap between both kinds of epistemic judgements and deontic obligations and possibilities is actually not too common. In their sample of 207 languages, van der Auwera & Ammann (2011a) found only 36 (or 17.4%) with markers that code both epistemic and situational necessity (i.e. epistemic deductive judgements and deontic obligations) and epistemic and situational possibility (i.e. epistemic speculative judgements and deontic permissions). The majority of these languages form a cluster in Europe and around the Mediterranean and do not occur at all in either of the Americas. It is more common, with 66 languages (or 31.9%), to allow some overlap, that is, to have markers that code either both epistemic and situational necessities, or that code both epistemic and situational possibilities, but not to have both kinds of overlap.

These languages are spread over the world. The biggest group of the sample, with 105 languages (or 50.7%) spread over the world but with a higher concentration in the Americas, lack any overlap between the two kinds of markers.

A fourth type of deontic modality is the **commissive**, where the speaker is certifying (or commits him/herself to ensure) that the action will take place. An example of a commissive would be something like *John shall have the book tomorrow*, where the speaker promises (is committing him/herself) that John will have the book tomorrow. A commissive may also be a threat, as in *You shall do as you're told or else*, where the speaker is promising (or threatening) that something will happen unless the addressee does as s/he is told.

8.5.3.2 *Dynamics*

While deontic modalities deal with actions initiated due to external factors, **dynamic** modalities deal with actions carried out due to internal factors.[140] These modalities can be divided into two subcategories, **volitives**, which denote willingness on the part of the subject to carry out an action, and **ability**, which denotes capacity on the part of the subject to carry out an action. English does not necessarily distinguish between the two formally; a sentence like *I can do it for you* may indicate both willingness and ability to do something. If distinction is necessary in English, *will* is used for willingness, as in *I will do it for you if I can*.

8.6 TMA in sign languages

It seems to be a universal that sign languages do not have grammatical tense. Instead time adverbials combined with the unmarked event sign serve to specify when an event occurred. Consider the following examples:

DGS (Sign Language: Germany)

(149) a. MUTTER WÄSCHE BÜGELN
mother laundry iron
'Mother is ironing the laundry.'

b. ENDLICH du-PRÜFUNG BESTANDEN
finally you-exam pass
'You've finally passed the exam.'

c. AUSBILDUNG BALD ANFANGEN
training soon start
'The training will start soon.'

(Adapted from *Die Brücke zur Welt der Gehörlosen* 2001: 21)

140. Termed 'agent-oriented' modalities in Bybee et al. (1994).

In each of the sentences in Example (149) the verb is in its base form, even though the events described take place at different places on the timeline. Unlike the English translations, there is no temporal information on any of the DGS verbs. In other words, unless an adverbial or some other contextual information is given, the location of the event on the timeline is not discernible.

An exception to this universal (or at the very least near-universal) might be Mexican Sign Language (LSM: Mexico), for which claims have been made of systematic past and future tense/aspect marking, while the present tense remains unmarked. In his study on the LSM tense and aspect system Fridman-Mintz (2005) shows that the perfective past is regularly marked with a head-lowering movement and a retention of the sign at its final location (the Hold segment), while the perfective future is marked with a head-lowering movement that starts from a Back Head posture (where the head is tilted backwards before moving forwards) combined with the final long Hold segment.[141] These tense distinctions are only found in the perfective aspect, however, and not in the imperfective, which is marked differently in its non-manuals (closed mouth with the corners down and eyebrows in a neutral position) from the perfective tense/aspect inflections and is tense neutral. Thus the sign EMPACAR[ipfv] 'pack' is tense neutral and may be translated as 'is/was/will be packing' depending on context (Fridman-Mintz 2005: 271).

Due to their perfect markers, ISL and ST might constitute two more exceptions to the tenselessness of sign languages, depending on whether one labels the perfect a tense or an aspect. The ST marker (glossed as HAP due to the mouth form) has two morphologically distinct forms, one for a positive utterance (150a) and one for a negative utterance (150b).

ST (Sign Language: Sweden)

(150) a. THIS ONE BOOK HAP READ
'He has read this book.'

_____*neg*

b. HAP-NOT INDEX-c OPEN WINDOW INDEX-c
'I have not opened the window.' (Bergman & Dahl 1994: 399, 401)

141. BSL shows a somewhat similar strategy in that a forward movement of head and shoulders may be used to indicate that event is located in the future, while a backwards movement of head and shoulders may be used to indicate that an event is located in the past. However, Sutton-Spence & Woll argue that this is not an instance of tense marking since "whole phrases and sentences are covered by these non-manual markers, not only verbs" (1999: 117). See also Jacobowitz & Stokoe (1988) for similar future marking strategies in ASL.

In (150) the perfect marker HAP indicates that the event took place before the reference point, but is still relevant at that point. In ISL the sign ALREADY functions in a similar way (Meir 1999). In both ISL and ST the markers indicate only that the event took place before a reference point (but is still relevant at that point); the reference point itself is relative and its location on the timeline is determined through context, such as various time adverbials.

While tense marking is virtually non-existent in the hitherto documented sign languages, aspect is very common indeed. A common strategy is to express the habitual or continuative aspects through reduplication.

Auslan (Sign Language: Australia)

(151) a. PRO-1 GO-TO+fast-red G-Y-M
'I go to the gym regularly.'
b. PRO-1 WAIT+fast-red PRO-2
'I have been waiting a long time for you.' (Johnston & Schembri 2009: 151)

In Example (151) a sense of duration or repetition is conveyed through the morphological process of reduplication. This strategy is found in numerous sign languages, for example ASL (Fischer 1973), ST (Ahlgren & Bergman 2006), and BSL (Sutton-Spence & Woll 1999), to mention only a few.[142]

Completive aspect is reported for IPSL, where the functional particle HO_GAYA indicates that an action or event has been completed (Zeshan 2003c: 164). It is possible that the ISL marker ALREADY could be analysed as a completive marker, as, for instance the auxiliary FINISH in Auslan (Johnston & Schembri 2009) might also be argued to be a completive marker, depending on such factors as obligatoriness.

Sign languages also tend to have various modal auxiliaries, such as those found in Auslan: CAN/CAN'T (two separate signs), MAY, SHOULD, WILL/WON'T (two separate signs), MUST and NEED (Johnston & Schembri 2009: 194) or those found in BSL: SHOULD, CAN, MUST, WILL (Sutton-Spence & Woll 1999). In BSL, as in DGS, the difference in obligation indicated between 'should' and 'must' is captured through speed and articulation. In BSL "SHOULD-ASK is smaller and less tense and strong than MUST-ASK" (Sutton-Spence & Woll 1999: 126), while in DGS SHOULD is signed more slowly and loosely than MUST (*Die Brücke zur Welt der Gehörlosen* 2001: 27).

Imperatives may either be expressed through functional particles or through morphological marking. In IPSL there are three functional particles expressing various imperatives: KARO 'neutral positive imperative', JA:O 'nonpolite positive imperative'

142. But see Bergman & Dahl (1994) for a discussion on why this is not a case of true aspect but rather one of ideophones.

and NAKARO 'negative imperative' (Zeshan 2003c:164). With morphologically marked imperatives in IPSL the sign for the desired action is made with a steady eye-gaze at the addressee and an optional "o" mouth form (mirroring the *-o* imperative suffix of spoken Hindi/Urdu). This is similar to the morphological imperative of ASL, where the sign of the desired action is made faster and sharper, with a direct eye-gaze at the addressee (Aikhenvald 2010:37).

8.7 Summary

Some very important verb phrase categories are tense, aspect and mood. Tense is the grammatical category that specifies where on a timeline an event is located. Absolute tenses relate the event to the moment of speech. Relative tenses relate the event to a given reference point. Languages may also code for remoteness or the degree to which the event is removed the reference point.

The perfect as a category does not fit neatly into either a tense paradigm or an aspect paradigm, as it locates an event on the timeline, but also specifies that the event is still relevant at the given reference point. Perfect in the past places the whole equation before a focus point and perfect in the future places the equation after a focus point.

Aspect is the grammatical category that specifies the perspective taken on an event. An event can be conceptualized as having a start and an end point, with a course between these points. An event may be viewed from within the start and end points, giving various imperfective aspects, or it may be viewed at or from without the start and end points, as a bounded whole, giving various perfective aspects. Aspect should not be confused with *Aktionsart* (actionality). While the former codes the perspective taken on the event, the latter codes the inherent semantics of the internal structure of the event.

Mode is the grammatical category that encodes the speaker's attitude toward a given proposition. The two major categories of mood are realis and irrealis. Realis codes that the event is a verifiable fact. Irrealis makes no claims as to the factuality of the proposition. The two major subcategories of modality are propositional modalities, which have to do with information, and event modalities, which have to do with action. Propositional modalities may be further subcategorized into evidentials, which code the kind of evidence a speaker has for a given proposition, and epistemics, which code the kinds of judgements a speaker is giving. Event modalities may either be deontic, where action is influenced through external factors, or they may be dynamic, where action is influenced through internal factors.

Sign languages typically lack the grammatical category of tense, and instead place events in time using various time adverbials. A few sign languages have a perfect marker. Aspect is commonly found in sign languages, especially habituals and progressives, which tend to be marked through reduplication. Sign languages also make use of various modal auxiliary verbs to code epistemic, deontic or directive modalities.

8.8 Keywords

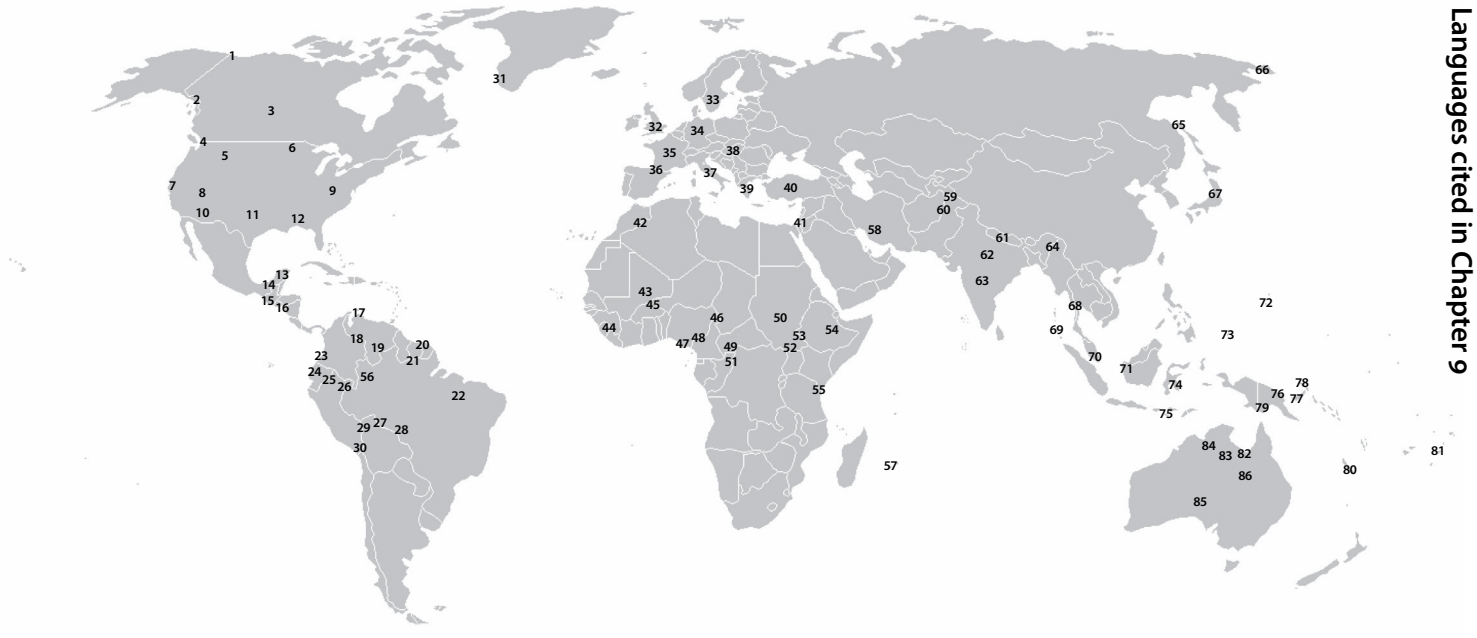
absolute/relative tense
Aktionsart
aspect
deontics
dynamics
epistemics
evidentials
mood/modality
perfective/imperfective
realis/irrealis
remoteness
tense
the perfect

8.9 Exercises

1. What is the difference between **absolute** and **relative tense**?
2. How does tense typically combine with **realis** and **irrealis**? Why?
3. What are **evidentials**? What type is most common?
4. How do sign languages pattern with respect to tense and aspect and how does this compare to spoken languages?
5. Is the following statement true or false? Motivate your answer.

   There is no difference between aspect and *Aktionsart*.

Languages cited in Chapter 9



1 Eskimo Pidgin
2 Tlingit
3 Cree, Plains
4 Halkomelem
5 Nez Perce
6 Michif
7 Pomo, Eastern
8 Chemehuevi
9 American Sign Language
10 Maricopa
11 Comanche
Kiowa
Wichita
12 Choctaw
13 Yucatec
14 Chol
15 Jakaltek
Tzutujil
16 Pipil
17 Papiamentu
18 Cuiba
19 Sanuma
20 Ndyuka
21 Tiriyo
22 Canela-Krahô
23 Awa Pit
Epena Pedee
24 Media Lengua
25 Záparo
26 Iquito
Yagua
27 Wari'
28 Aikaná
Kanoé
29 Araona
30 Aymara
31 Greenlandic, West
32 English
33 Swedish
34 German
35 French
36 Basque
37 Italian
38 Hungarian
39 Greek Sign Language
40 Turkish
41 Al-Sayyid Bedouin Sign Language
Israeli Sign Language
42 Berber, Middle Atlas
43 Koyraboro Senni
44 Kisi
45 Koromfe
46 Hdi
47 Igbo
48 Babungo
49 Sango
50 Krongo
51 Lingala
52 Päri
53 Murle
54 Oromo, Harar
55 Ma'a/Mbugu
56 Nadëb
57 Réunion Creole
58 Persian, Old
59 Rushan
60 Pashai
61 Chantyal
Chepang
Kham
62 Hindi
Indo-Pakistani Sign Language
63 Marathi
64 Nocte
65 Nivkh
66 Chukchi
67 Japanese
68 Karen, Pwo
69 Nicobarese, Car
70 Semelai
Singapore Bazaar Malay
71 Mualang
72 Chamorro
73 Palauan
74 Tukang Besi
75 Kambera
76 Mauwake
77 Bali-Vitu
78 Kuot
79 Kala Lagaw Ya, Saibai
80 Bislama
81 Fijian
82 Kayardild
Yukulta
83 Wambaya
84 Gurindji Kriol
85 Pitjantjatjara
86 Kalkatungu

Chapter 9

Simple clauses

Clauses need to have a certain number of participants in order to be grammatically acceptable. The number of necessary participants is determined by the type of verb in the clause. This number may then be modified through different strategies. The participants of a clause are marked in various ways in the different types of clauses. This chapter discusses some basic issues on clause participants and valency. Section 9.1 brings up the core semantic, pragmatic and syntactic roles participants may have in a clause, and discusses the types of argument alignments found in the world and indexing of arguments on the verb. Section 9.2 deals with valency, the number of obligatory participants a clause must have, and how valency may be adjusted in various ways. In 9.3 I bring up some issues related to simple clauses in sign languages.

9.1 Clause participants

A sentence will consist minimally of a **predicate** and the **arguments** of that predicate to form a proposition.[143] Typically, the predicate minimally consists of a verb, which is the crucial element that determines the structure of the proposition. The arguments are typically realized by noun phrases, which fill the structural slots laid out by the verb in the proposition. For example, in the sentence *John threw the ball*, there is a verb (signifying the act of throwing) and two arguments, one entity that does the throwing (*John*) and one entity that gets thrown (*the ball*). Verbs differ in how many arguments they require. A verb like *throw* needs two arguments, while a verb like *run* only needs one; while *John ran* is perfectly acceptable and the destination is not obligatory to specify, a sentence like ??*John threw* would only work in extremely specific contexts and would normally require that we specify what got thrown. A verb like *give* normally requires three arguments, the entity giving something, that which is being given, and the recipient of the gift, as in *John gave Mary a ball*.

The participants that are necessary to make a sentence grammatical can be called **core participants** (or **arguments**). They are the subject, direct and indirect objects, depending on what the verb requires. Sentences may also contain participants that are not obligatory for the sentence to be grammatical, **peripheral participants** (or **adjuncts**).

---

143. Below we will see that predicates may also be without arguments (zero arguments).

For example, *John ran* contains one core participant (argument), the subject (*John*), while *John ran to the bus stop* contains one core participant (*John*) and one peripheral participant (*to the bus stop*). In *John gave Mary a ball* we have the three core participants subject (*John*), direct object (*a ball*) and indirect object (*Mary*), while in *John gave Mary a ball for her birthday* we also have a peripheral participant (*for her birthday*). The sentences would not be grammatical without their core participants, whereas they are still grammatical if we leave out the peripheral participant; compare *___ *ran* (core participant missing) with *John ran* ___ (peripheral participant missing) and *___ *gave Mary a ball* / **John gave Mary* ___ / ??*John gave* ___ *a ball* (a core participant missing) with *John gave Mary a ball* ___ (peripheral participant missing). For more on core participants (arguments) and peripheral participants (adjuncts), see 9.2 below.

In all the sentences above the noun phrases have various semantic roles and grammatical relations. For instance, *John* is consistently an active participant of the act (he runs, throws, gives) and is also consistently the subject of each sentence, while *the ball* is consistently a passive participant of the act (it gets thrown and gets given) and is also consistently an object of the sentence. However, if we say *The ball was kicked by John*, then *the ball* becomes the subject even though it retains its role as a passive participant of the proposition. In the same way, the semantic role of *the candle* as an instrumental stays the same in the two sentences *Mary lit the room with a candle* and *The candle lit the room*, even though is it an object in the first sentence and a subject in the second. This shows that semantic roles and grammatical relations do not necessarily overlap and that, in fact, it makes sense to view the two as separate linguistic domains that may interact in various ways. Semantic roles specify the underlying relationship that arguments have with their verbs in terms of how they develop the meaning of the clause or sentence. Grammatical relations specify the relationship between the arguments and the verb in terms of the syntactic function that they carry in the clause or sentence. Furthermore, the arguments of a proposition can have different pragmatic roles, depending on how they reflect the flow of the information given in the clause or sentence. The following sections will give brief definitions of some general terms and concepts that will serve as a background for the discussion in this and the following chapters. For a thorough overview of the various functions of the noun phrase in the clause, see Andrews (2007b).

9.1.1 Semantic roles

As mentioned, **semantic roles** (also called **thematic roles** or **theta** (*θ*) **roles**) denote the roles that participants play in a given situation. It is important to keep in mind that semantic roles do not denote any inherent properties of the participants but that they specify what role the participants have in relation to the verb. Languages differ as to how they may code various semantic roles or how they may be grouped together.

In principle one could argue that a semantic role is a conceptual notion and that the roles of participants in each situation are unique and therefore subtly different from the roles in any other situation, which would yield an infinite amount of semantic roles. This is, strictly speaking, true, but also in a sense irrelevant and, above all, not helpful for linguistic descriptions. Of all the infinite situations available to us, there are some common features that can be used as a basis for delimiting a smaller number of typical semantic roles that tend to occur in languages and tend to be grammatically marked in some way. This section will give examples of some common semantic roles but makes no claims whatsoever of being an exhaustive list of all potential semantic roles that can be found across languages. It is important to bear in mind that there are no absolute boundaries between the labels, but rather that they constitute more or less well-defined clusters along a continuum of such semantic notions as control and action, where, for example, the cluster on one end typically denotes entities with a high degree of control that instigate action and the cluster on the other end typically denotes entities that have a low degree of control and do not wilfully engage in any action. In the sentence *John kicked the ball*, for example, *John* is an entity with a high degree of control instigating action while *the ball* is an entity with a low degree of control undergoing that action.

A prototypical **agent** is the entity that performs an action and has a high degree of control. An example of an agent would be *Mary* in *Mary opened the door with the key*. There is often an element of volition (that the act has been carried out on purpose), although that would be lacking in non-conscious entities such as natural forces, as in *The wind blew open the door*, where it is questionable whether *The wind* could be argued to act through volition. In some cases it might thus be justified to distinguish between agent and **force**, where the latter is a non-conscious instigator of action.

A prototypical **instrument** is an entity which is used to perform an action. Very often an agent uses an instrument as a tool to carry out the action, as in *Mary opened the door with the key*, where *the key* is the instrument. Notice, however, that in the sentence *The key opened the door* the instrument (*the key*) remains the same even though its grammatical relation is now that of subject. The instrument typically does not have conscious control, but is merely a tool being used for an action.

A prototypical **experiencer** is an entity getting cognitive stimulus one way or another. An example of an experiencer would be *the cat* in *The cat saw the mouse*. The difference between an experiencer and an agent is one of control and volition: while an agent typically consciously performs an action, the experiencer typically does not have control of the cognitive stimulus that it receives. Compare the sentences *Mary saw the mouse* and *Mary looked at the mouse*. In the former Mary became aware of the mouse through the sense of sight but may or may not have wished to become aware of it. In the latter Mary chooses to look at the mouse. Thus, in the former sentence,

Mary is rather an experiencer than an agent, while in the latter she is rather an agent than an experiencer.

A prototypical **recipient** is an entity receiving something. An example of a recipient would be *John* in the sentence *Mary gave the book to John*. The role of recipient is not entirely overlapping with **benefactive**, which prototypically is an entity that benefits from an action (but might not receive anything), such as *I hung the laundry for Mary*, where Mary is the benefactive.

A prototypical **patient** has very little or no control and volition and is the entity affected by the action. An example of a patient would be *the door* in the sentence *Mary opened the door*. Other examples would be *The door opened* (no control, no volition), *Mary fell* (no control, no volition), *John pushed Mary* (*Mary* is the patient with minimal control), and so on.

A prototypical **theme** is an entity which changes location or to which a location is assigned. An example of a sentence with a theme would be *Mary gave the book to John*, where *the book* is the theme.

The above-mentioned semantic roles are often expressed as subjects, objects or indirect objects (see below). The following are more often expressed as adverbials or obliques (optional arguments or participants).

A prototypical **comitative** is the entity accompanying someone or something, as in *John went to the beach with the dog*, where *(with) the dog* is the comitative. A prototypical **purpose** (or **purposive**) is an entity which is the reason for an action, as in *John went out for the newspaper*, where *(for) the newspaper* is the purpose.

A prototypical **locative** is the entity expressing the point in space for an event or entity, as in *Mary opened the door in the hall*, where *(in) the hall* is the locative. A prototypical **goal** is the entity expressing the end point for a motion, as in *John took the bus to school*, where *(to) school* is the goal, while a prototypical **source** is the origin of a motion, as in *Mary came in from the balcony*, where *(from) the balcony* is the source.

A prototypical **temporal** is a noun phrase expressing the point in time for an action or event, as in *John went out for groceries at noon/in the morning/at ten o'clock*, where *(at) noon/(in) the morning/(at) ten o'clock* express the temporal semantic roles.

9.1.2 Pragmatic roles

We must always keep in mind that language is used in a social context of speakers, addressees, and people or events far removed either in time or place; in spontaneous utterances or carefully prepared narratives or formalized rituals or anything in between. These social contexts include shared sociocultural knowledge and other shared information, such as in terms of previous speech or other kinds of background knowledge. In other words, languages are very rarely (if ever) employed to just randomly throw out sentences. Rather, humans use language in a structured manner,

maximizing the coherence of the message or information they are trying to get across. Pragmatics deals with how this practical use of language manifests itself linguistically. For thorough discussions on discourse strategies, see, for example, Kärkkäinen et al. (2007) and Gumperz (1982), the latter of which is a classic by now. For a very thorough overview on pragmatics in general, see Levinson (1983), another classic piece of work. A very accessible introduction to pragmatics for the beginner is Yule (1996). Chapter 12 gives a further discussion of speech acts and similar topics.

Very simplified, **pragmatic roles** describe the status of the information given in a sentence or clause. Languages have various ways of presenting essentially the same information from different perspectives. For instance, in the two sentences *Bill opened the door* and *The door was opened by Bill*, the same basic information is given, namely that Bill engaged in the act of opening the door. The difference lies in how that information is presented, or packaged. The first sentence is mainly about Bill and what he did, while the second sentence is mainly about the door and what happened to it. Pragmatic roles thus denote how the sentence is structured in terms of information status and information packaging. For a very thorough discussion on information packaging, see Foley (2007). For a detailed discussion on pragmatic roles in general, see Lambrecht (1994).

The **topic** (or **theme**) is essentially the entity about which something is said, i.e. what the sentence is about. The rest of the sentence is a **comment** on the topic. Thus the topic in *Bill opened the door* is *Bill* while *opened the door* is the comment. Similarly, the topic in *The door was opened by Bill* is *the door*, while *was opened by Bill* is the comment. Prototypically, the topic represents old or already known information that does not need further highlighting. In other words, the topic "prototypically denotes a presupposed established entity" (Foley 2007: 405) in a context which constitutes shared information between the speaker and the addressee. Topics tend to be grammatical subjects, which in a sense is rather straightforward, since subjects tend to be the central participants in a clause (see below). This should not be taken to imply that topics and subjects are the same categories, even though there is a high degree of overlap. For a thorough discussion on the pragmatic role of topic and its interplay with the grammatical relations of subject and object, see Lambrecht (1994: 131ff) and Dalrymple & Nikolaeva (2011) with further references.

A sentence can have more than one topic. The **primary topic** is the more important entity, while the **secondary topic** is the entity that stands in some relationship with the primary topic. The utterance as a whole is about this relationship (Nikolaeva 2001). Consider the following:

(152) a. Whatever became of Jenny?
b. She married Peter,
c. and she's still deeply in love with him.

In (152b) we have a comment (*married Peter*) on the topic *she*, which refers to *Jenny*. In (152c) we have two topics, *she* (i.e. Jenny) and *him* (i.e. Peter) and the utterance is about the relationship between these two entities. The sentence is primarily about Jenny, thus making her the primary topic, but it also adds information about Peter, making him a secondary topic. Givón (e.g. 1984 and 2001a and previous) has argued that the primary topic is the origin of the grammatical relation *subject* while the secondary topic is the origin of the grammatical relation *direct object*. For more on grammatical relations, see 9.1.3 below.

The **focus** (or **rheme**) constitutes new, not previously established (or at least presupposed not to be previously established) information about the topic. In other words, the focus constitutes information that is added to the pool of shared knowledge. It can be part of the comment, or even constitute the whole comment. In fact, an entire clause can function as focus, if it consists entirely of new information. This is easiest to illustrate by means of question and answer dialogues, where the focus is underlined:

Q: Who opened the door?
A: Bill opened the door.
Q: What did Bill open?
A: He opened the door.
Q: What happened?
A: Bill opened the door.

Notice that in a real-life dialogue the answers could have consisted of only the focus, while that information which is shared between the speaker and the addressee does not have to be repeated in order to get across the relevant new information:

Q: Who opened the door?
A: Bill.
Q: What did Bill open?
A: The door.
Q: What happened?
A: Bill opened the door.

What is crucial to keep in mind here is that information packaging may take various shapes, and consist of different pragmatic roles, depending on a multitude of factors, such as the flow of a narrative and how much information can be presupposed to be shared.

Pragmatic roles are very commonly marked through intonation. In English, for example, a focused element tends to be stressed, as in *Who opened the door? **Bill** opened the door*, where Bill is the focus (with the stress indicated in bold). Other ways of pragmatically marking an element may be through **fronting**, where the focused

element is placed at the beginning of the sentence. Compare, for example, the sentences *I like roses* with *Roses, I like*. In the latter sentence the focused element (*roses*) has been placed at the very beginning of the sentence, it has been fronted. Yet another way of pragmatically marking an element is through a **cleft construction**, where, very simplified, focus is achieved by expressing the sentence as a relative clause in which a noun phrase (NP_i) and the relativized NP are coreferential (see 11.2.3 for more on relative clauses). In English this expressed as NP_i be [… NP_i…]$_{Relative\ clause}$, as in *It is roses that I like*, where *It* is NP_i and *roses* the relativized NP which is coreferential with NP_i. The NP_i is typically referred to as the **clefted constituent** and is commonly, though not always, placed furthest to the left of the clause.

9.1.3 Grammatical relations and alignment

Grammatical relations (or **syntactic roles**) are formal categories that signal the syntactic function an argument has in the clause or sentence. In other words, they are formal categories for expressing what function a participant has in a given situation or proposition. Languages tend to have up to three distinct grammatical relations, traditionally termed subject, object (or direct object) and indirect object, the occurrence of which is largely dependent on the transitivity of the verb. In subsequent sections we will see that these are not adequate categories to capture the various core grammatical functions found in languages of the world, which are better captured by the three core semantico-syntactic functions S, A and P.[144] For an accessible overview on grammatical relations from different theoretical perspectives, see Farrell (2005). For thorough discussions on grammatical relations typology, see, for example, Givón (2001a: 173ff), Andrews (2007b), Dryer (2007a) and Bickel (2010). For a classic on the definition of the subject, see Keenan (1976).

9.1.3.1 *Subjects and objects*

Grammatical relations tend to have a set of formal properties. The **overt coding properties** refer to the "perceptually discernible features of the grammatical code such as morphology, intonation and word order" (Givón 2001a: 175). The overt coding properties typically relevant for grammatical relations are word order, nominal morphology and verb agreement. The **behaviour-and-control** properties refer to the syntactic processes that are governed by grammatical relations. In principle this refers to any process that answers *Yes* to the question "[d]oes one have to mention the subject and/or direct-object

---

144. Dixon (e.g. 1979 and 1994; cf. also Foley & Van Valin 1984) has argued that the notion of **syntactic pivot** better captures why languages group participants in a certain way. The syntactic pivot is, put very simply, the argument around which the rest of the sentence revolves. In many languages (especially those with a nominative-accusative alignment, see 9.1.3.3), the subject is the syntactic pivot of the sentence.

G[rammatical] R[elation] in describing the grammatical behaviour of a particular construction?" (Givón 2001a: 178). An example of a behaviour-and-control process is passivization, which can be described as a construction where the direct object is promoted to the subject of the clause, as in *John threw the* ***ball*** (direct object) passivized into *The* ***ball*** (subject) *was thrown*. (For more on passive constructions, see 9.2.2.1 below.)

The **subject** is the central, most prominent, noun phrase in the clause and traditionally defined as the "doer" of an action, or the entity which carries out the action expressed by the verb. While this may overlap with the semantic role of agent, the two are not identical. Recall, for example, that different semantic roles may fill the subject position (underlined), as in ***The man*** (agent) *opened the door with a key*, ***The key*** (instrument) *opened the door*, and ***The door*** (patient) *opened*; in the three sentences an agent, an instrument, and a patient respectively fill the grammatical slot of the subject. In English the position of a participant in a clause with respect to the other participants signals which grammatical relation the participant has. In *Mary phoned John*, for example, we know that the NP *Mary* is the subject (Mary is the "doer") because of its placement in the clause: subjects come before the verb in English. For more on how different languages order the participants in the clause, see Chapter 10. Nominal morphology (especially case marking, see 7.1.3) may also be used to signal the grammatical relation of a participant. In English this is evident only in the pronominal system: in the sentence *She phoned John* we know that *she* is the subject of the clause because of the morphological form of the NP. Something like **Her phoned John* would not be acceptable. In languages with case systems the subject is typically in the nominative or ergative case. Very often the verb carries overt markers indexing such properties as person and number of the subject, known as **verb agreement** (see below, 9.1.3.5). In English, for instance, the verb agrees with the subject, but no other participant: in *The man shoots ducks* the verb agrees in person and number with the subject (3rd person singular *-s*). Something like **The man shoot ducks*, where the verb agrees with the object (*ducks*), is not possible.

The **object** is basically the core argument of the verb which is not a subject and is traditionally defined as the "recipient" or "undergoer" of an action. Again, while this may overlap with various semantic roles, the two are not identical. And again, the formal coding properties of word order, nominal morphology and verb agreement signal the grammatical relation of the argument. In English, for example, the object comes after the verb: in *Mary phoned John* we know that the NP *John* is the object ("recipient" or "undergoer") because of its placement in the clause. Nominal morphology may also mark objects; in English this is evident with pronouns, where something like *Mary phoned him* is acceptable but **Mary phoned he* is not. In languages with case systems the object is typically in the accusative, ablative or dative case.

With verbs that demand two objects a distinction is often made between direct and indirect objects. A **direct object** is traditionally defined as the "undergoer"

or patient or theme of the action, while the **indirect object** is traditionally defined as the "recipient" or "addressee" of the action. For instance, the verb *sell* requires that someone does the selling (subject) of something that gets sold (direct object) to someone (indirect object). Most commonly languages make a formal distinction between direct and indirect objects. In English the indirect may be formally differentiated from the direct object, as in *John gave the book to Mary*, there the preposition *to* indicates that Mary is the indirect object. However a sentence like *John gave Mary the book*, where no formal differentiation is made, is equally acceptable. Languages with case systems often, but by no means always, use different cases for direct and indirect objects.

9.1.3.2 *Transitivity*

The above discussion has consistently referred to arguments needed by the verb and the number of participants in a clause. **Transitivity** is the category that specifies whether or not a verb can take an object. Verbs fall into different classes, the major divide being between those that can take object(s) and those that cannot. Thus an **intransitive** (which literally means "not transitive") verb has only one participant, the subject. An example of a transitive verb would be *sleep*, as in *John/The man slept*, where *John* or *The man* is the only participant (subject). Something like **John slept something* is not acceptable. A **transitive** (from Latin *transire* 'to carry over') verb, on the other hand, has two participants that stand in some kind of relationship to each other, typically in that the subject acts on the object somehow (an act is "carried over" from the subject to the object). An example of a transitive verb would be *kill*, as in *The man killed a fly*, where *The man* is the subject and *a fly* is the object. Something like **The man killed* is not acceptable (or at the very least, would only be acceptable in highly specific contexts). A **ditransitive** (literally "two-transitive") verb takes two objects, which means that the clause has three participants in total. An example of a ditransitive verb would be *sell*, as in *John sold the house to Mary*, where *John* is the subject, *the house* is the direct object, and *Mary* is the indirect object.

9.1.3.3 *Subject alignment*

The discussion above shows that one way of defining subjects is that they are the sole arguments of intransitive verbs. It might seem straightforward to go further and define subjects as not only the sole argument of intransitive verbs, but also the primary argument of transitive verbs. After all, the subjects, in English, look the same in both cases:

(153) a. He went.
S V
b. He saw him.
S V O

In (153) the subjects formally look the same even though the first sentence is intransitive and the second transitive. The form of the object is never possible to use in a subject position; something like **Him went* is not acceptable. Clearly, then, the subject seems to occupy the slot of the primary argument of any verb. But now consider the following example:

Pitjantjatjara (Australian (Pama-Nyungan): Australia)

(154) a. minyma-ngku tjitji nya-ngu
woman-ERG child-Ø(ABS) see-PAST
S O V
'The woman saw the child.'

b. tjitji a-nu
child-Ø(ABS) go-PAST
S V
'The child went.'

c. tjitji-ngku minyma nya-ngu
child-ERG woman-Ø(ABS) see-PAST
S O V
'The child saw the woman.' (Bowe 1990:10)

In (154a and c) we have transitive clauses with the two arguments subject (*minymangku* 'woman.ERGative' and *tjitjingku* 'child.ERGative' respectively) and object (*tjitji* 'child' and *minyma* 'woman' respectively), while in (154b) we have an intransitive clause with only one argument, the subject (*tjitji* 'child'). Notice, however, that the subjects do not always look formally the same: in the transitive clauses the subjects consistently have the ergative case ending *-ngku* but in the intransitive clause the subject has no ending at all (it is in the absolutive case). In fact, the form of the subject in the intransitive clause is identical with the form of the objects in the transitive clauses!

Because languages have different systems of grouping (**aligning**) the core arguments of intransitive and transitive clauses, it makes more sense to differentiate these three arguments as follows: the subject (S) is the single argument of an intransitive clause – the position occupied by *He* in (153b) and *tjitji* in (154b); *minyamangku* and *tjitjingku* in (154a) and (154c) respectively are the most agent-like arguments of the clause and as such may be labelled agents (A); while the position occupied by *him* in (153b) and *tjitji* and *minyma* in (154a and c) respectively are the most patient-like arguments of the clause and as such may be labelled patients (P).[145] In other words, in order to be able to better capture what is going on, we make use of terms relating both to grammatical relations (*subject*) and to semantic roles (*agent* and *patient*) to

145. While it is more common to use the symbol P for the argument occupying the most patient-like slot of a transitive clause, the symbol O (for object) is sometimes also used, as in Dixon (1994).

describe the semantico-syntactic alignment of arguments in different languages; we may think of them as **participant roles**.

There are five logical ways of aligning S, A and P, the core arguments of intransitive and transitive clauses, as illustrated in Figure 9.1.

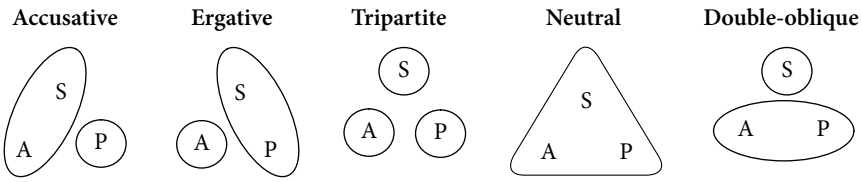


Figure 9.1 The five logical possibilities of aligning the arguments S, A and P.

In a **nominative-accusative** (or **accusative**) system, S and A are grouped together in that they are marked in the same way, while P is marked differently. The example above demonstrates how English pronouns show an accusative alignment. Hungarian is an example of a language where full noun phrases have an accusative system, with the S and A both in the nominative case (*-Ø*) and P in the accusative (*-et*):

Hungarian (Uralic (Ugric): Hungary)

(155) a. a lány-Ø áll
the girl-Ø stand.INDEFOC.3SG
S
'The girl is standing.'

b. a lány-Ø ir-ja a level-et
the girl-Ø write-DEFOC.3SG the letter-ACC
A P
'The girl is writing the letter.' (Kenesei et al. 1998: 195)

In Comrie's (2011a) sample of 190 languages, 52 (27.4%) are accusative, spread all over the world except Southeast Asia and Papua New Guinea. Most commonly accusative languages either have overt marking for both the nominative and the accusative or an overt marker for the accusative only (as in the Hungarian example above). However in six languages (3.2% of the entire sample) in the sample the nominative is overtly marked while the accusative is unmarked. Murle is an example of such a language:

Murle (Nilo-Saharan (Surmic): Sudan)

(156) a. adokony ɛɛt-i
runs man-NOM
S
'The man runs.'

(156) b. agam kayuu-wi kulugit-Ø
catches eagle-NOM fish-Ø(ACC)
A P
'The eagle catches a fish.' (Arensen 1982: 50, 139)

In (156) both S and A have overt nominative marking (*-(w)i*) while P is in the unmarked accusative. This is still an accusative system, however, since the crucial fact is that S and A are treated the same way while P is treated differently. Since full noun phrases and pronouns might behave differently in terms of alignment (as is the case with English, for instance, compare Example (153) with the discussion on neutral systems below), Comrie (2011b) treated pronouns in a separate survey. Of 172 languages, 64 (37.2%) have an accusative system. Three (1.7% of the entire sample), Aymara (Aymaran (Aymaran): Bolivia), Igbo and Maricopa, have a marked nominative and an unmarked accusative, as they do with full noun phrases. Murle and Harar Oromo have a marked nominative for full noun phrases, but a standard accusative system for pronouns, and Middle Atlas Berber has a marked nominative for full noun phrases but a neutral system (see below) for pronouns.

In an **ergative-absolutive** (or **ergative**) system S and P are marked in the same way while A is marked differently. Pitjantjatjara, as shown above, is an example of a language with an ergative system. Epena Pedee is an example of a language where both the full noun phrases and the pronouns have an ergative system.

Epena Pedee (Choco (Choco): Colombia)

(157) a. josé-Ø kʰāi-hí
PN-Ø sleep-PAST
S
'José slept.'

b. josé-pa pʰáta-Ø kʰo-hí
PN-ERG plantain-Ø eat-PAST
A P
'José ate (the) plantain.'

c. mí-Ø kʰāi-itʰée
1SG-Ø sleep-FUT
S
'I will sleep.'

d. mí-a pʰáta-Ø kʰo-hí
1SG-ERG plantain-Ø eat-PAST
A P
'I ate (the) plantain.'

e. mí-Ø pee-wǻ-da a-hí na imamá-pa
1SG-Ø kill-PROG-DECL say-PAST this tiger-ERG
P A
'"This tiger is killing me!" he said.' (Harms 1994: 9, 22, 45, 69, 87)

In (157a and b) we see that the S is unmarked, just as the P is (*josé-Ø* and *pháta-Ø* respectively) while the A is treated differently, marked with the ergative suffix (*josé-pa*). The personal pronoun *mí* 'I' behaves in the same way, remaining unmarked in the S and P slots (Examples (c) and (e)) while carrying the ergative suffix in the A slot (Example (d)). In Comrie's two surveys (2011a and b) 32 of 190 languages (16.8%) have an ergative system for full noun phrases and 20 of 172 languages (11.6%) have an ergative system for their pronouns. While languages with an ergative system in Comrie's samples are spread globally, they are completely absent in Africa, and almost completely absent in Europe and on the northern part of the Eurasian landmass (Basque and Chukchi being the respective exceptions).

In a **tripartite** system S, A and P are all marked differently. This is very rare cross-linguistically and is found in only four languages (2.1%) in Comrie's (2011a) survey on full noun phrases: Hindi, Marathi (Indo-European (Indic): India), Nez Perce (Penutian (Sahaptian): USA) and Semelai (Austro-Asiatic (Aslian): Malaysia). All but Marathi also have a tripartite system for the pronouns (while the pronouns in Marathi have an accusative system). In Nez Perce, for example, the S is unmarked while the A carries an ergative suffix (*-nim*/*-nm*/*-m*) and the P carries a direct object suffix (*-ne*), as shown in Example (158).

Nez Perce (Penutian (Sahaptian): USA)

(158) a. kaa wáago hi-x̣íic'em-ne háama-Ø
and now 3.NOM-be.angry-PFV man-Ø
S
'And now the man became angry.'

b. 'iceyéeye-nm x̣áx̣aas-na hi-náas-wapci'yaw-na
coyote-ERG grizzly-DO 3NOM-PLDO-kill-PFV
A P
'Coyote killed the grizzlies.' (Rude 1985: 83, 88)

The most common system in Comrie's (2011a and 2011b) samples is the **neutral** system, where S, A and P all look the same. It is generally found in languages with little or no morphological marking. English is an example of a neutral system in the full noun phrases (though not with pronouns, see above): the NP carries no overt marking to indicate what syntactic function it has in the clause. Instead, that is determined by the word order. For instance, in the sentences *John* (S) *slept* and *John* (A) *called Bill* (P), the arguments are all unmarked. Another example of a language with a neutral system is Koromfe, where again all the arguments are unmarked.

Koromfe (Niger-Congo (Gur): Burkina Faso, Mali)

(159) a. a hem koŋ gɔl
ART water(SG) DET.NHUM.SG boil
S
'The water is boiling.'

b. bʌdini homs a hem koŋ
PN heat ART water(SG) DET.NHUM.SG
A P
'Badini boils the water.' (Rennison 1997: 268)

In Comrie's sample mapped for full noun phrases, 98 of 190 (or 51.6%) languages have a neutral system, while in the sample mapped for pronouns 79 languages (or 45.9%) have a neutral system. These languages are spread all over the world, but concentrate in the areas where languages have little or no morphological marking or where most of the morphological marking occurs on the verb.

The fifth logical possibility of subject argument alignment is that A and P look the same while S looks different, called **double-oblique** (or sometimes "**accusative-focus**", cf. for example, Whaley 1997: 158). This is exceedingly rare and is only known to occur in a few Iranian languages of the Pamir region, specifically with Rushan (Indo-European (Iranian): Tajikistan) pronouns in clauses of the past tenses, although there have been reports of a double-oblique system for both full noun phrases and pronouns in Pashai (Indo-European (Iranian): Afghanistan) (Skalmowski 1974).

Rushan (Indo-European (Iranian): Tajikistan)

(160) a. mu tā wunṭ
1SG.OBL 2SG.OBL see.PAST
A P
'I saw you.'

b. tā mu wunt
2SG.OBL 1SG.OBL see.PAST
A P
'You saw me.'

c. az-um sut
1SG.ABS-1SG go.PAST(M.SG)[146]
S
'I went.' (Payne 1980: 156, 158)

In (160a and b) the pronouns (*mu* 'I, me' and *tā* 'you' respectively) look exactly the same irrespective of whether they function as A arguments or P arguments. In other

146. Intransitive verb stems show gender and number alternations in the various tenses that refer to the past.

words, A and P are ambiguous here. However, when the pronoun functions as an S argument, i.e. as the sole argument of an intransitive clause, it takes a different form, namely *az*. Furthermore, the verb agreement suffix only appears on the S, here the first person marker *-um*. This double-oblique system is in fact reflecting an intermediate stage in an ongoing shift from an ergative system to an accusative system.[147]

The languages in APiCS present a rather different picture, as shown in Table 9.1.

Table 9.1 Comparison between WALS[148] and APiCS languages for subject alignment. Adapted from Comrie (2011a and 2011b) and Michaelis et al. (2013: features 58, 59). Absolute numbers in parentheses.

		Full NPs				**Pronouns**			
	Value	**WALS**		**APiCS**[149]		**WALS**		**APiCS**	
1.	Neutral	51.1%	(96)	81.6%	(62)	45.3%	(77)	31.6%	(24)
2.	Nominative/accusative (standard)	24.4%	(46)	17.1%	(13)	35.9%	(61)	68.4%	(52)
3.	Nominative/accusative (marked nominative)	3.2%	(6)		(0)	1.8%	(3)		(0)
4.	Ergative/absolutive	17%	(32)	1.3%	(1)	11.8%	(20)		(0)
5.	Tripartite	2.1%	(4)		(0)	1.8%	(3)		(0)
6.	Active/inactive[150]	2.1%	(4)		(0)	1.8%	(3)		(0)
	Total		188		76		170		76

It thus seems that pidgins and creoles are much less likely to have an ergative system than non-creoles, and that pidgins and creoles are more likely than non-creoles to have a neutral system for their full noun phrases but an accusative system for their pronouns.

---

147. The actual path of development is quite a bit more complicated than that: the Old Persian accusative system developed into an ergative system when the passive particle got reanalysed as an active verb. This ergative system has now largely been lost again in favour of the accusative system due to language contact, but various intermediate stages in the Pamir languages have led to this double-oblique system, which is still partly found in Rushan. See further Payne (1980 and 1989). For an accessible overview of the intense and complex contact situation involved in the various stages of the Iranian languages, see Utas (2009). For more on contact-induced language change, see 13.2.

148. For the purpose of comparison between APiCS and WALS patterns, the creole languages in the WALS sample have been taken out, which is why the figure for WALS differ slightly here from those earlier in the text.

149. The language for value 4 is Gurindji Kriol (Mixed language (Gurindji, Kriol): Australia).

150. This refers to split systems (see 9.1.3.6 below). The added value 'none' in Comrie's survey on pronoun alignment has been ignored here. The languages for this value are Wari', Wichita and Canela-Krahô (Macro-Ge (Ge-Kaingang): Brazil).

9.1.3.4 *Object alignment*

Languages may also differ in how they align the objects of a ditransitive clause. In a ditransitive clause we have three arguments, the subject, the direct object and the indirect object. Or, more exactly, we have three participant roles: the agent-like argument (A) discussed above, the recipient-like argument (R) and the theme-like argument (T). An example of a ditransitive construction is:

(161)	John	gave	Mary	a flower
	A	R		T

In (161) *John* is the agent-like argument, *Mary* is the recipient-like argument and *a flowe*r is the theme-like argument. For a detailed and very accessible overview of ditransitive constructions, see Malchukov et al. (2010a) with references, which serves as the introduction to Malchukov et al. (2010b), where a collection of studies on ditransitive constructions in different languages can be found.

Languages differ in whether and how they encode the R and T arguments of a ditransitive construction compared to the P argument of a monotransitive construction. There are five logical ways of aligning the R and T arguments, as illustrated in Figure 9.2:

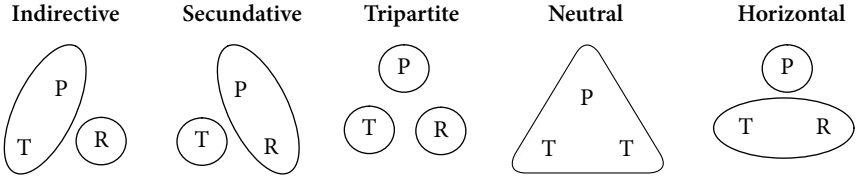


Figure 9.2 The five logical possibilities of aligning the arguments P, T and R.

In an **indirective** alignment the T of a ditransitive verb is treated in the same way as the P of a monotransitive verb. German is an example of a language with indirective alignment:

German (Indo-European (Germanic): Germany)

(162) a.	Der	Mann	kaufte	den	Ball
	ART.NOM.SG	man	bought	ART.ACC.SG	ball
		A			P

'The man bought the ball.'

b.	Der	Mann	gab	dem	Kind	den	Ball
	ART.NOM.SG	man	gave	ART.DAT.SG	child	ART.ACC.SG	ball
		A		R			T

'The man gave the child the ball.' (source: personal knowledge)

In the ditransitive construction in (162b) the T (*den Ball* 'the ball') is encoded in the same way as the P of the monotransitive construction in (162a) (*den Ball* 'the ball'). This is the most common strategy in Haspelmath's (2011a) sample, where 189 of 378 languages (or 50%) have an indirect alignment.

In a **secundative** alignment it is the R of a ditransitive verb that is encoded in the same way as the P of a monotransitive verb, as in Kham:

Kham (Sino-Tibetan (Bodic): Nepal)

(163) a. ŋa:-Ø no:-lai ŋa-rɨ̃:h-ke
1SG-NOM 3SG-OBJ 1SG-see-PFV
A P
'I saw him.'

b. ŋa:-lai bəhtanji y-ā:-ke-o
1SG-OBJ potato give-1SG-PFV-3SG
R T A
'He gave me a potato.' (Watters 2002: 67f)

In (163b) the recipient-like argument (*ŋa:lai*) of the ditransitive construction carries the object marker *-lai* just as the patient-like argument of the monotransitive construction in (163a) does (*no:lai*). This alignment strategy is found in 66 (17.5%) languages in Haspelmath's sample.

In a **neutral** alignment (also called **double-object** alignment) P, R and T are all encoded the same way:

Hdi (Afro-Asiatic (Biu-Mandara): Nigeria)

(164) a. ŋ̀ghà-n-ŋ̀gh-í tá krì
see-3-see-1SG OBJ dog
A P
'I saw a dog.'

b. vlá-n-vl-í tá xə̀n tá kóɓù
give-3-give-1SG OBJ 3PL OBJ money
A R T
'I give them money.' (Frajzyngier & Shay 2002: 180, 182)

In (164) the R and the T (*xə̀n* 'them' and *kóɓù* 'money' respectively) of the ditransitive construction in (b) are both marked with the object marker *tá*, as is the P (*krì*) of the monotransitive construction in (a). This alignment strategy is found in 83 languages (22%) of Haspelmath's (2011a) sample.

In a **tripartite** alignment P, T and R are all encoded differently. This is very rare (Malchukov et al. 2010a), but can be found in Kayardild:

Kayardild (Australian (Tangkic): Australia)

(165) a. ngarrka-ngku ka yankirri luwa-rni
man-ERG PRES emu.ABS shoot-NPST
A P
'The man is shooting the emu.'

b. maku dun-maru-tha wuu-ja nguku-wuru
woman.NOM spouse-VD-ACT give-ACT water-PROP
A R T
'A woman gives water to her spouse.' (Evans 1995: 139, 336)

In (165) the P (*yankirri* 'emu') of the monotransitive construction in (a) is marked in the ablative case, while in the ditransitive construction in (b) the R (*dunmarutha* 'to (her) spouse') is marked in the verbal dative case and the T (*ngukuwuru* 'water') is marked in the proprietive case. All three arguments are thus encoded differently.

The fifth logical possibility, where the R and the T of the ditransitive clause are encoded the same way but the P of the monotransitive clause is encoded differently, has so far not been reported (Malchukov et al. 2010a).

The smallest group in Haspelmath's (2011a) sample consists of 40 languages (10.6%) that have a mixed system, such as English. Consider, for example, the sentences *John gave a book to Mary* and *John gave Mary a book*. In the former sentence there is a formal distinction between the R (*to Mary*) and the T (*book*), where in the latter there is not (*Mary* and *book* respectively). This strategy is usually called **dative shift** (see further 9.2.3.2 below).

9.1.3.5 *Verb agreement*

Reference to the participants in a clause may also be coded on the verb. Many languages have various morphological ways of marking participant reference on the verb, **verb agreement** (sometimes also **verb coding** or (verb) **concord**). What makes this a case of agreement is that the form of the marking (whether an affix or a stem change) is chosen paradigmatically according to the semantic properties of the participant referred to. While English has very little morphological marking, the verb *be* changes form according to the number and person properties of the participant referred to. Thus a first person singular participant reference requires the form *am* (*I am*), while a first person plural participant reference requires the form *are* (*we are*), and so on. Similarly, one of the few morphological markings in English is the present tense suffix *-s*, which is required if the S/A ('subject') argument is in third person singular, as in *The girl sees a flower*. If the S/A changes to the plural, the verb loses the suffix, as in *The girls see a flower* (but not **The girls sees a flower*).

Languages differ in whether or not they allow verb agreement to be the only reference to a participant in a clause. In English, for example, a free-form reference to the

argument is obligatory. A sentence like **am* is not possible, even though the form of the verb clearly indicates that the participant referred to is the first person singular. This is sometimes called **grammatical agreement** as opposed to **anaphoric agreement**, where the verbal reference to a participant is the only needed reference to the argument in question. In Italian, for example, the word *parlo* 'I speak' forms a perfectly grammatical sentence. The verb ending *-o* indicates that the participant referred to is in the first person singular and the free form *io* 'I' is not obligatory. Languages that allow anaphoric agreement are sometimes called **pro-drop** languages, while languages that do not allow anaphoric agreement are sometimes called **lexical argument** languages. It is much more common to not require a subject pronoun for a clause (i.e. to allow 'pro-drop'). In Dryer's (2011c) sample of 711 languages, 437 (61.5%) languages, spread all over the world, have anaphoric agreement, as is the case in Italian, while only 82 (or 11.5%), found especially in Europe and West Africa, require the pronominal subject to be present, as is the case in English. Even if we add Dryer's fourth value, which indicates the 67 (9.4%) languages where "pronominal subjects are expressed by subject pronouns that occur in a different syntactic position from full noun phrase subjects" (Dryer 2011c), we get 149 languages (21%), which is still a minority. However, it makes sense to differentiate between the two values, since included in this latter value are languages where the subject pronoun is a portmanteau for both pronominal and TMA values. This latter value also includes languages where the subject pronoun is obligatory even if a full noun phrase is expressed in the clause. In both these cases one could argue that the pronouns in question are fulfilling similar tasks as verbal affixes, even though they are phonologically free words, and that as such they could be argued to belong to the verbal complex in the same way as verbal affixes do. An example of a language where this is particularly evident is Bali-Vitu, where a preverbal particle expresses both TMA values and person values.

Bali-Vitu (Austronesian (Oceanic): Papua New Guinea)

(166) e Lingei te vigi-ho na kalabaka
ART PN REAL.3 hit-2SG PREP stick
'Lingei hit you (SG) with a stick.' (Ross 2002a: 374)

In (166) the preverbal particle *te* is a portmanteau for realis mood and third person, and it is obligatory even though the full noun phrase (*Lingei*) is present. Languages of this kind are predominantly found in sub-Saharan Africa and the western Pacific.

In 32 languages (4.5%) in Dryer's sample the pronominal subject is expressed by a clitic which can have variable hosts while in 61 languages (8.6%) the presence of the subject pronoun is optional. Chemehuevi is an example of a language with subject pronoun clitics, where the pronominal subject attaches to the first word of the clause unless the clause contains a demonstrative, in which case it attaches to the demonstrative.

Chemehuevi (Uto-Aztecan (Numic): USA)

(167) a. wahaku-a-n totoci-vɨ punikai-vɨ
two-OBL-1SG head-PL(OBL) see-PAST
'I saw two heads.'

b. aipaci aŋa-ja-n kwipa-vɨ
boy.OBL that-OBL-1SG hit-PAST
'I hit that boy.' (Press 1975: 109, 186)

In (167a) the first person singular marker *-n* attaches to the first word in the sentence, the numeral *wahaku-* 'two', while in (167b) it attaches to the demonstrative *aŋa-* 'that'. Recall that a clitic is a phonetically bound unit that is not constrained by the syntax of its host. Languages with subject pronoun clitics are predominantly found in North America and Australia. An example of a language where the pronoun subject is optional in the clause is Chantyal, where overt marking of the participant is not necessary if it is otherwise clear from the context.

Chantyal (Sino-Tibetan (Bodic): Nepal)

(168) mənchi-sə thar-ra bənnu lɦi-si-rə sar-ji chala
person-ERG mountain.goat-DAT gun hit-ANT-SEQ kill-PFV skin
tar-si-rə, ənnarabɦūri thuti-si-rə, tɦem-əŋ kɦa-si-ə,
extract-ANT-SEQ viscera pull.out-ANT-SEQ house-LOC bring-ANT-SEQ
dewri-ri pəkə-ysi-rə, ca-i
pot-LOC cook-ANT-SEQ eat-PFV
'The man shot a mountain goat and killed it. He skinned it, cleaned it, took it home, put it in a pot, cooked it and ate it.' (Noonan 2003: 334)

In (168) there is no overt subject pronoun nor is there any verb affix referring to any of the arguments; once the participants have been introduced (the man and the mountain goat), there is no need to repeat any reference to them, since it is clear from the contexts who is doing what to whom. Compare that to the English translation, where the pronouns are obligatory (*he*, the subject, is used once and *it*, the object, is repeated seven times). Languages where the subject pronoun is optional are predominantly found in East and Southeast Asia as well as Australia.

The remaining 32 languages (4.5%) in Dryer's sample have a mixed system where presence or absence of the pronominal subject in the clause is either dependent on various factors such as person, animacy, transitivity, and so on, or where more than one of the above-mentioned the strategies are allowed.

The figures for the WALS sample are only minimally affected by the subtraction of the two contact languages in Dryer's (2011c) sample: Ndyuka, which is coded as having obligatory pronouns in subject position, and Sango, which is coded as having anaphoric agreement (the latter differs from the APiCS coding; see below). The

pattern for whether or not the subject pronoun is obligatory differs radically in the APiCS sample, as shown in Table 9.2.

Table 9.2 Comparison between WALS and APiCS languages for expression of pronominal subject. Adapted from Dryer (2011c) and Michaelis et al. (2013: feature 62). Absolute numbers in parentheses.

	Value	WALS		APiCS[151]	
1.	Pronominal S obligatory	11.4%	(81)	65.8%	(50)
2.	S-affixes on verbs (anaphoric agreement)	61.5%	(436)	5.3%	(4)
3.	Pronominal S clitics	4.5%	(32)		(0)
4.	Pronominal S in different position	9%	(64)	2.6%	(2)
5.	Pronominal S optional	8.6%	(61)	22.4%	(17)
6.	Mixed	4.5%	(32)	3.9%	(3)
	Total		709		76

This means that while non-creoles are much more likely to express the pronominal subject through verbal affixes, pidgins and creoles seem much more likely to have obligatorily expressed subject pronouns (i.e. they do not seem likely to allow 'pro-drop').

Languages also differ in how many and which arguments are marked on the verb. Those of us mostly used to Western European languages might assume that verbal affixes only say anything about the 'subject', the A argument in transitive clauses. However, it is actually much more common to mark both the A and the P of transitive clauses on the verb. In Siewierska's (2011d) sample of 378 languages, only 73 (or 19.3%) restrict their marking to the A only, as is the case in English. This is particularly common in Eurasia and eastern Africa but is not found at all in the Australian languages of the sample. 193 languages (51.1%), however, mark both the A and P on the verb. This is a very common strategy in all areas of the world except Eurasia and is particularly dominant in North America, New Guinea and Australia. An example of a language that marks both the A and the P is Palauan:

Palauan (Austronesian (Palauan): Palau)

(169) te-'illebed-ii a bilis a rengalek
3PL(A)-hit-3SG(P) ART dog ART children
'The children hit the dog.' (Georgopoulos 1991: 30)

151. The languages for value 2 are Lingala, Ma'a/Mbugu (Mixed language (Bantu, Cushitic): Tanzania), Media Lengua and Michif; the languages for value 4 are Eskimo Pidgin and Papiamentu (Creole (Spanish-lexified): Netherlands Antilles); the languages for value 6 are Bislama (Creole (English-lexified): Vanuatu), Réunion Creole (Creole (French-lexified): Réunion) and Singapore Bazaar Malay. In APiCS Sango is coded as having optional subject pronouns.

In (169) the prefix *te-* indicates that the 'subject' (A) is a third person plural (namely *rengalek* 'the children'), while the suffix *-ii* indicates that the 'object' (P) is a third person singular (namely *bilis* 'dog'). Much less common is to mark only the P on the verb. This is found in only 24 (6.3%) languages in Siewierska's sample, mainly in Africa and South America. Kisi is an example of such a language:

Kisi (Niger-Congo (Southern Atlantic): Guinea)

(170) í sílá-ŋ lé
1SG advise-3SG NEG
'I didn't advise him.' (Childs 1995: 72)

In (170) the verbal suffix *-ŋ* denotes the P, while the A (*í* 'I') appears as a free form.

A very rare strategy indeed is to mark either the A or the P, but never both at the same time. This is found only in six languages (1.6% of the sample), three of which – Awa Pit (Barbacoan (Barbacoan): Colombia), Iquito (Zaparoan (Zaparoan): Peru) and Tiriyo (Example (171)) – are located in northern South America, while the others are scattered, with Kiowa (Kiowa Tanoan (Kiowa Tanoan): USA) in North America, Kalkatungu (Australian (Pama-Nyungan)) in Australia and Chepang (Sino-Tibetan (Bodic): Nepal) in South Asia.

Tiriyo (Cariban (Cariban): Suriname)

(171) a. manko j-ennoo-ne makapa-pona
1:mother 1P-send-PAST.PFV Macapá-DIRC
'My mother sent me to Macapá.'
b. ji-nmuku wi-po-ka-e
1-son 1A-clothes-PRIV.VZR:PRES.IPFV-CTY
'I am undressing my son.' (Meira 1999: 263, 270)

In Tiriyo the verb may only index for one argument, the choice of which is based on person ranking. Thus if one argument is either first or second person (1/2) and the other argument is third person (3), the 1/2 person is indexed on the verb irrespective of whether it constitutes an A or a P, as shown in Example (171). In the first sentence the P (*j-* 'I') is indexed, since it refers to the first person, while the A (*manko* 'my. mother') is the third person. In the second sentence the A is indexed, because this time the A is the first person (*wi-* 'I') while the P is the third (*nmuku* 'son'). If both arguments are either the first and/or the second person the "speech act participants only"[152] prefix (*k-*/*kï-*) is used and if both participants are the third person the "non-speech act participants" prefix (*n-*/*ni-*/*nï-*) is used:

152. Speech act participants (SAP) means the first or second persons, i.e. the speaker(s) or addressee(s).

(172) a. k-ëta
1/2A/P-hear.PAST
'I heard you.'/'You heard me.'
b. n-apëi
3A/P-catch:PRES.PFV
'S/he/it has caught (it/him/her).' (Meira 1999: 284, 289)

Roughly one fifth (82 languages or 21.7%) of Siewierska's (2011d) sample lack verbal person marking altogether, mainly found in West Africa, the Caucasus and South and Southeast Asia. Babungo, for example, lacks verbal person marking.

Babungo (Niger-Congo (Bantoid): Cameroon)

(173) mə jìa wèe
1SG hold.PRES child
'I hold a child.' (Schaub 1985: 226)

Languages also differ with respect to the alignment of their verbal marking. Siewierska (2011a) mapped 380 languages for their verbal person marking alignment. In contrast to noun phrases and pronouns, where the most common system is a neutral one (cf. Comrie 2011a and 2011b as well as the previous section), this is found only for about a fifth (84 languages or 22.1%) of Siewierska's sample for verbal person marking, corresponding to those languages that do not have verbal person marking (and consequently found in the same areas, see above). Instead, the most common system is the accusative alignment, found in 212 languages (55.8%) spread over all areas of the world. Pipil is an example of a language with accusative verb person marking alignment:

Pipil (Uto-Aztecan (Aztecan): El Salvador)

(174) a. ni-panu
1SG-pass
'I pass/cross.'
b. ti-panu
2SG-pass
'You pass/cross.'
c. ni-mits-ita-k
1SG-2SG-see-PAST
'I saw you.'
d. ti-nech-ita-k
2SG-1SG-see-PAST
'You saw me.' (Campbell 1985: 68)

In (174) the S of the intransitive clauses (*ni-* 'I', *ti-* 'you') are marked in the same way as the A in the transitive clauses (*ni-* 'I', *ti-* 'you'), while the P are marked differently (*-mits-* 'you', *-nech-* 'me').

All other systems are comparatively rare in Siewierska's sample. The ergative system is found only in 19 languages (5%), mainly in South America and Southeast Asia. An example of a language with ergative verbal person marking alignment is Tzutujil:

Tzutujil (Mayan (Mayan): Guatemala)

(175) a. x-in-war-i
compl-1sg.abs-sleep-IV.phrf
'I slept.'

b. x-at-war-i
compl-2sg.abs-sleep-IV.phrf
'You slept.'

c. x-in-aa-choy
compl-1sg.abs-2sg.erg-hit
'You cut me.'

d. x-at-nuu-ch'ey
compl-2sg.abs-1sg.erg-hit
'I cut you.' (Dayley 1981: 85, 88)

In (175) the S (*-in-* 'I' and *-at-* 'you') look the same as the P (*-in-* 'me' and *-at-* 'you') while the A look different (*-nuu-* 'I' and *-aa-* 'you').

Slightly more common than the ergative system is the so-called active system of verbal person marking alignment, found in 26 languages (6.8%) in Siewierska's sample. Here S aligns with either A or P depending on various semantic or pragmatic factors. In Yagua, for example, the S may look like the P in various discourse contexts.

Yagua (Peba-Yaguan (Peba-Yaguan): Peru)

(176) a. sa-jú́ú́y Anita
3sg-fall PN
'Anita falls.'

b. sa-suuta Rospita raruvą́ą́́va-níí
3sg-wash PN down.river-3sg
'Rospita washes him/her downriver.'

c. múúy jú́ú́-níí munuñúmiy
there fall-3sg savage
'There falls the savage.' (Payne 1985a: 45, 47, 57)

The S in Example (176a) looks like the A in (176b) (*sa-* '3sg'), while in (176c) the S looks like the P in (176b) (*-níí* '3sg'). A few languages, 11 (2.9%) in Siewierska's sample, arrange the alignment according to a referent hierarchy and usually the argument with the higher ranking gets the special treatment.

The tripartite system is very rare indeed and is not included in the map. An example of a language with the tripartite system for the second person singular as well as the first person singular and plural is Yukulta.

Yukulta (Australian (Tangkic): Australia)

(177) a. waranaŋkuḷu-ka-ti
go.NEG.DES-1SG.S-PRES
'I'm trying to go.'

b. ṭalmata-ŋa-nti ŋita
chop.IND-1SG.A-FUT wood
'I'll chop the wood.'

c. tʸinkaka-nki ŋata
follow.IMP-1SG.P me
'Follow me.' (Siewierska 2011a citing Keen 1983: 215, 237, 239)

In (177) each of the verbal person markings for the S (*-ka-* 'I'), A (*-ŋa-* 'I') and P (*-nki* 'me') arguments look different. Languages may also have more than one type of system, split systems, found in 28 languages (7.4%) of Siewierska's sample.

9.1.3.6 *Split systems*

Languages tend to group together items with similar roles, so that those elements that have similar kinds of functions also end up looking formally similar (this is what Haiman 1980b calls **isomorphism**). Thus there are probably functional motivations for grouping S, A and P in various ways. What ties S and A together, functionally, are the semantic roles of AGENT and the pragmatic roles of TOPIC. For example, in Mary (S) *ran ahead* and Mary (A) *batted the ball*, both the S and A are the agents and topics of the clause. What ties S and P together, on the other hand, are the semantic notions of change of state and the pragmatic roles of FOCUS. For example, in ***The lake*** (S) *froze* and *Mary opened* ***the envelope*** (P), both S and P undergo a change of state and constitute the focus of the clause. In languages with split systems, the motivation for grouping S with either A or P lies with various semantic and pragmatic factors. For instance, languages may have various alignments depending on the semantic and/or pragmatic properties of the intransitive clause (split intransitivity) while others may have various alignments depending on the semantic and/or pragmatic properties of the transitive clause and more specifically the semantic and pragmatic roles that the arguments get in the clause (split ergativity).

With **split intransitivity** the S is marked differently depending on the semantic role it has. If it is essentially an agent-like role, the S will be marked as an A of a transitive clause, showing accusative alignment. If, however, the S has an essentially patient-like role, it will be marked as the P of a transitive clause, showing ergative alignment.

The two types of S arguments may be termed S_A (for the type that aligns with the A of a transitive clause; S_A arguments are sometimes called 'actors') and S_P (for the type that aligns with the P of a transitive clause; S_P arguments are sometimes called 'undergoers'). The semantic role of S may be determined by the semantic properties of the verb; for instance, that dynamic verbs take an S_A while stative verbs take an S_P. Or the S may be split along the lines of volition, where S_A is used when the argument is volitional and S_P when the argument is not. Verbs that specifically take S_A arguments are sometimes called 'unergative' verbs while verbs that specifically take S_P arguments are sometimes called 'unaccusative' verbs. Chochtaw is an example of a language with split intransitivity, where dynamic verbs take 'actor' S arguments (S_A) while stative verbs take 'undergoer' S arguments (S_P).

Choctaw (Muskogean (Muskogean): USA)

(178) a. hilha-li-tok
dance-1SG.NOM(ACTOR)-PAST
'I danced.' [S = +VOLITIONAL]

b. sa-hohchafo-h
1SG.ACC(UNDERGOER)-hungry-PRED
'I am hungry.' [S = −VOLITIONAL]

c. chi-bashli-li-tok
2SG.ACC(ACTOR)-cut-1SG.NOM(UNDERGOER)-PAST
'I cut you.' [A = +VOLITIONAL; P = −VOLITIONAL]

d. is-sa-sso-tok
2SG.NOM(ACTOR)-1SG.ACC(UNDERGOER)-hit-PAST
'You hit me.' [A = +VOLITIONAL; P = −VOLITIONAL]

e. sa-hohchafo-cha tobi hoponi-li-tok
1SG.ACC(UNDERGOER)-hungry-SS bean cook-1SG.NOM(ACTOR)-PAST
'I was hungry, so I cooked some beans.' (Davies 1986: 14–5, 28)

In Example (178a) the intransitive clause contains a dynamic verb with a volitional argument and the S takes the same form as in the A in the transitive clause in (178c) (*-li-* 'I'). Here the S and A have in common that they are both actors in the clause. In the intransitive clause containing a stative verb with a nonvolitional argument (b), the S takes the same form as the P in the transitive clause in (178d) (*-sa-* 'I, me'). Here the S and P have in common that they are both undergoers in the clause. In (178e) the same person is once an undergoer (S_P) and once an actor (S_A). In Eastern Pomo (Hokan (Pomoan): USA) the semantic feature of volition is particularly dominant in governing the choice of the alignment. The S_A is used when the intransitive argument is volitional, while the S_P is used when it is nonvolitional. This

works both ways: with some verbs either S_A or S_P can be used, affecting the meaning of the whole clause.

Eastern Pomo (Hokan (Pomoan): USA)

(179) a. há mi·pal šá·k'a
1SG.A 3SG.O killed
'I killed him.'

b. xá·su·là wí ko·k^{h}óya
rattlesnake 1SG.O bit
'(A) rattlesnake bit me.'

c. há ĉe̊·xélka
1SG.A(S_A) slide/slip
'I am sliding.' (on purpose)

d. wí ĉe̊·xélka
1SG.O(S_P) slide/slip
'I am slipping.' (not on purpose) (McLendon 1978: 1–3)

In Example (179) the intransitive takes an S_A argument if the subject acts with volition, while it takes an S_P argument if there is minimal or no volition in the act. Notice that the verb in (179c and d) is the same, but gets different readings ('sliding' versus 'slipping') depending on the argument chosen, and the choice of the argument itself depends on the presence or absence of volition with the subject.

FOURSOME SPLIT

The Saibai dialect of Kalaw Lagaw Ya (Australian (Pama-Nyungan): Australia) has four different alignments depending on the type of NP:

i. tripartite for singular pronouns
ii. neutral for plural pronouns
iii. accusative for proper names and kin terms
iv. ergative for common nouns (Comrie 1981a: 8ff)

Split intransitivity may be based on other criteria, such as animacy or pragmatic roles. The Yagua Example (176) above, for instance, illustrates a split based on pragmatic roles. The S_A and the S_P in the example are both nonvolitional arguments, but the S_P expresses a change of scene and is thus a focussed argument.

While split intransitivity is based on the semantics and/or pragmatics of the intransitive clause, **split ergativity** is based on the semantics and/or pragmatics of

the transitive clause.[153] One determining factor of the split in alignment may be the animacy of the arguments, especially the agent-like arguments. Thus it is often the case that pronouns, which are higher up the animacy hierarchy (or empathy hierarchy, see, for example DeLancey 1981), have an accusative alignment, while all other nominals have an ergative alignment. This is sometimes called agent-worthiness or topic-worthiness (cf., for example, Payne 1997). In Kham, for instance, the speech act participants (SAP) have an accusative alignment while all other nominals (i.e. those ranking lower than the speaker and the addressee on the animacy hierarchy) have an ergative alignment.

Kham (Sino-Tibetan (Bodic): Nepal)

(180) a. la:-Ø si-ke
leopard-ABS die-PFV
'The leopard died.'

b. Tipəlkya-e la:-Ø səih-ke-o
PN-ERG leopard-ABS kill-PFV-3SG
'Tipalkya killed a leopard.'

c. no:-ye la:-Ø səih-ke-o
3SG-ERG leopard-ABS kill-PFV-3SG
'He killed a leopard.'

d. ŋa:-Ø la:-Ø ŋa-səih-ke
1SG-NOM leopard-ABS 1SG-kill-PFV
'I killed a leopard.'

e. nɨ-Ø ŋa-lai nə-rɨ̄:h-na-ke
2SG-NOM 1SG-OBJV 2SG-see-1SG-PFV
'You saw me.' (Watters 2002: 66–8)

In Example (180a, b and c) the S and the P are marked in the same way, with the absolutive (-Ø) case, while the A are marked with the ergative (*-(y)e*) case. In (180d) the A argument, an SAP, is marked with the nominative (-Ø) case, while the P argument, not an SAP, is marked with the absolutive (-Ø) case. In (180e), however, where both participants are SAPs, we see a classic accusative system, where the A is marked in the nominative (-Ø) case, while the P is marked in the 'objective' (*-lai*) case.

---

153. There is no particular reason why this should be called "split ergativity" and not, for example, "split accusativity". As Scott DeLancy says, "[i]t is very telling that we talk about 'ergativity' – and have conferences and workshops on topics like 'Ergativity in Amazonian Languages' – but conferences or thematic volumes devoted to 'accusativity' or 'nominativity' (which I think is the preferable term) seem to be a less prominent feature of the linguistic landscape. (…) Our tradition inclines us to think of 'ergativity' as some special deviation from the norm; in fact it is a non-category, it is simply the absence of nominativity" (2006: 6). See also McGregor (2009).

Another determining factor of the split alignment may be the tense, mood or aspect of a clause. More often than not the tenses referring to the past and perfective or completive aspects trigger an ergative alignment while nonpast tenses and imperfective aspects trigger an accusative alignment. Chol is an example of a language where the accusative alignment is used only in imperative declarative sentences while all others follow an ergative alignment.

Chol (Mayan (Mayan): Mexico)

(181) a. ƈa-h k'el-e-Ø
ASP-1SG.A see-PFV-3SG.P
'I saw it.' (perfective)

b. ƈa til-ig-on
ASP come-PFV-1SG.SP
'I came.' (perfective)

c. mi-h wahl-en-et
ASP-1SG.A mock-IPFV-2SG.P
'I ridicule you.' (imperfective)

d. mi-h suht-el
ASP-1SG.SA return.IPFV
'I return.' (imperfective) (Quizar & Knowles-Berry 1988:77–8)

In (181) the perfective sentences (a and b) display an ergative alignment, where the single argument of the intransitive clause is expressed with the suffixes used for P arguments in transitive clauses. In the imperfective sentences (c and d), however, the single argument of the intransitive clause is expressed in the same way as A arguments of transitive clauses.

9.2 Valency

We have seen that verbs differ in how many (if any) objects they can take. Verbs also differ in their **valency**, that is, in how many arguments they obligatorily have to take.[154] The former, transitivity, overlaps a great deal with valency, but the two should not be confused. A ditransitive verb, for example, takes two objects (direct and indirect),

154. Strictly speaking there are two different notions, **semantic valency**, which refers to the amount of necessary participants to a verb, and **syntactic valency**, which refers to the amount of arguments that are present in the clause. While the semantic valency remains the same – *put* is always a trivalent verb (i.e. it has a valency of three) – the syntactic valency may vary depending on whether and which valency-adjusting operations are used. See further below. For a thorough discussion on the typology of valency, see Kulikov (2010).

as in *John gave the book to Mary*, where *the book* is the direct object and *Mary* is the indirect object. But consider a verb like *place*, where somebody (subject) places something (object) somewhere, as in *John placed the book on the table*. Here we have only one object (*the book*); *on the table* is not an indirect object. It is, however, a necessary argument to the verb *place*. Sentences like **John placed the book*, **Placed the book on the table* or **John placed on the table*, where one of the arguments is missing, are not acceptable. An example of a **monovalent verb** (i.e. a verb with a valency of one) is *dance*, since only one participant is necessary to make the clause grammatical: *Mary danced*. An example of a **bivalent verb** (a verb with a valency of two) is *kill*, since two participants are necessary to make the clause grammatical: *John killed the fly* (but not **John killed* or **Killed the fly*).

In English there is no such thing as an **avalent verb** (i.e. a verb with a valency of zero, also called **zero-intransitives** or **ambient clauses**), so every clause needs a subject, even if it does not refer to any actual entity, a so-called **dummy** or **expletive subject**. An example of a dummy subject can be found in the sentence *It is raining*, where *It* does not refer to any actual entity. Avalent verbs usually involve environmental conditions, and are especially typical for weather conditions (cf. Dryer 2007a: 267ff). In Italian, for example, there is no dummy subject.

Italian (Indo-European (Romance): Italy)

(182) piove
rain.3SG.PRES
'It is raining.' (source: personal knowledge)

In (182) the verb is inflected for the third person singular, but there is no noun phrase argument in the clause. Notice that avalent clauses are not the same as clauses with 'pro-drop', since with 'pro-drop' the pronoun may optionally be included or left out, the choice hinging on various factors such as, for example, focus. A sentence like *lui parla* 'he speak.3SG.PRES' is as grammatical as *parla* '(he) speak.3SG.PRES'. This is, however, not possible with weather expressions: something like **lui piove* is not acceptable.

Since having a dummy subject in weather expressions is the norm for Germanic languages, which many of us are familiar with, especially English, it might come as a surprise that this is, in fact, extremely rare cross-linguistically. In a genetically balanced survey of 218 languages, I found that only six (2.8%) languages, including English,[155] had a dummy subject in weather expressions, as shown in Map 9.1 below.

155. The other five languages are Comanche, Cuiba (Guahiban (Guahiban): Colombia), Car Nicobarese (Austro-Asiatic (Nicobarese): Nicobar Islands), Tlingit (Na-Dene (Tlingit): USA) and Záparo (Zaparoan (Zaparoan): Ecuador).

The most common strategy for expressing 'it is raining' (or some similar environmental expression) is to have a referential subject, found in 128 languages (58.7%). An example of a referential subject construction can be found in Mauwake, where the clause contains both a full noun ('rain') and a verb ('descend'):

Mauwake (Trans-New Guinea (Madang): Papua New Guinea)

(183) ipia or-om-ik-eya
rain descend-SS.SIM-be-2/3SG.DS
'It was raining.' / 'It rained.' (Berghäll 2010: 143, 237)

Avalent constructions, as in the Italian example above, are found in 78 languages (35.8%). The remaining six languages[156] (2.8%) have a noun used predicatively in the clause, as in Kambera, where the noun 'rain' may be used to express 'it is raining' (or, '(There is) rain').

Kambera (Austronesian (Central Malayo-Polynesian): Indonesia)

(184) urang
rain(NOUN)
'It is raining.' / '(There is) rain.' (Klamer 1998: 410)

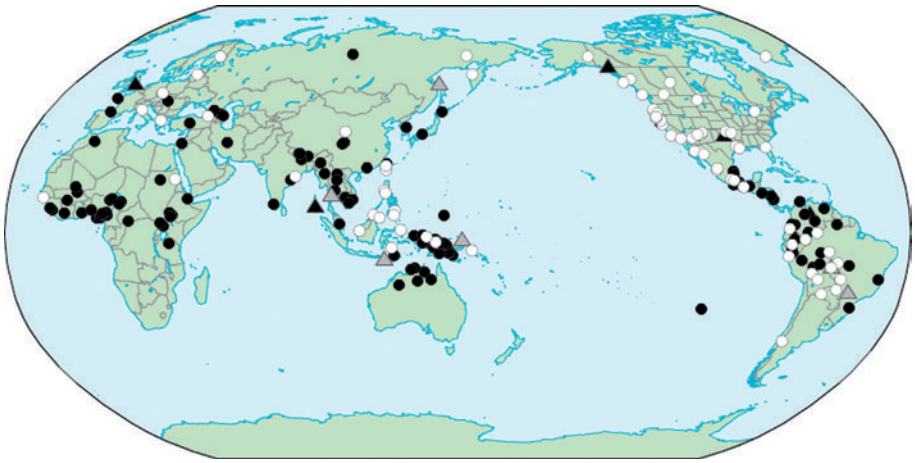
As Map 9.1 below shows, there is some areal patterning to the strategies: while the few languages with a dummy subject and with referential subjects are scattered about, avalent clause strategies cluster in Southern and Eastern Europe, on the Southeast Asian islands and in North and South America but are almost completely absent in Africa, Mainland Asia and Meso-America, while the referential subject strategy clusters in Africa, mainland Asia, Australia and Papua New Guinea, as well as Meso and South America, but is almost completely absent in insular Southeast Asia and North America.

Whereas arguments are obligatory, **adjuncts** are not, but instead constitute optional noun phrases included in a clause. An example of an adjunct can be found in *Mary danced on the street*, where *on the street* is an adjunct (a non-obligatory noun phrase); a sentence like *Mary danced* is perfectly acceptable without any further specification.

All languages have various strategies for adjusting the valency of a clause. Valency may be increased, decreased, or the roles of the arguments may be swapped. While it is beyond the scope of this section to give an exhaustive list of the various valency alternating strategies known across languages, some examples of the more commonly discussed strategies will be mentioned below.

---

156. These are: Aikanã, Nivkh (Isolate: Russia), Kambera, Kanoé (Kapixana (Kapixana): Brazil), Pwo Karen (Sino-Tibetan (Karen): Myanmar) and Kuot (Isolate: Papua New Guinea).



Map 9.1 Survey of weather expressions. Black triangles: dummy subject (6 languages); grey triangles: predicative noun (6 languages); black dots: referential subject (128 languages); white dots: avalent construction (78 languages). For a full legend, see http://dx.doi.org/10.1075/z.176.additional.

9.2.1 Increasing valency

Increasing the valency of the clause involves bringing in a new necessary participant into the clause. Merely adding optional information, such as adjuncts, does not entail valency increase. The crucial point is that with the valency increasing device the new participant becomes an argument, an obligatory participant, without which the clause would be ungrammatical.

9.2.1.1 *Causatives*

Causative constructions essentially merge two separate events into one single complex event which in turn increases the valency of the original event by one. For example, in the clause *Mary laughed* we have one argument. A causative construction of the basic event (*Mary laughed*) would be something like *John made Mary laugh*. Here two separate events are merged into one: the causing event, where John (the **causer**) does something, and the caused event, where Mary (the **causee**) does something (*laughs*) because of John's action. It is the notion of causation that demands an additional participant: someone has to be the causer of the event (a sentence like **made Mary laugh* is not acceptable, for example). Causative constructions have been the focus of much research and interest, and it is quite beyond the scope of this section to adequately deal with the full complexity of the topic. For accessible discussions on causatives, see Comrie (1989:165ff) and Dixon (2000). A very thorough discussion indeed on causatives is Song (1996).

Causatives are commonly divided into three different types, lexical, morphological and analytic causatives. In **lexical causatives** the semantics of the verb itself contains a notion of causation. Compare the difference between *die* and *kill*, where the former does not contain any notion of causation (*John died* does not imply any other actor in the event) but the latter does (in *Peter killed John* we may assume that Peter caused John to die). It is extremely common, if not universal, for languages to have at least some lexical causatives. In **morphological causatives** a morphological process is applied to the base verb in order to get the notion of causation. Very often the morphological process involves affixation, but it may also involve other morphological processes that the language makes use of, for example various non-linear markers such as tone, ablaut or the root-and-pattern. Turkish is an example of a language with morphological causative constructions.

Turkish (Altaic (Turkic): Turkey)

(185) a. Hasan koş-tu
PN run-PAST
'Hasan ran.'

b. (ben) Hasan-ı koş-tur-du-m
1SG PN-ACC run-CAUS-PAST-1SG
'I made Hasan run.'

c. Hasan kitab-ı oku-du
PN book-ACC read-PAST
'Hasan read the book.'

d. (ben) Hasan-a kitab-ı oku-t-tu-m
1SG PN-DAT book-ACC read-CAUS-PAST-1SG
'I made Hasan read the book.' (Kornfilt 2003: 331)

In (185) the intransitive clause *Hasan ran* (a) is changed into a transitive clause *I made Hasan run* (b) by means of the causative suffix *-tur*. Notice that Hasan goes from subject to direct object, which seems straightforward enough, since the causer must be the subject, which makes the causee an object. In (185d), however, where the transitive clause of (185c) has been made ditransitive, Hasan goes from subject to indirect object, which might seem counter-intuitive. Here we have an instance of what may be termed a 'causee accessibility hierarchy' of subject > direct object > indirect object > oblique (Comrie 1989: 176), where the causee (in this case Hasan) moves to the highest available slot in the causative clause (i.e. the left-most available slot). The slot of subject will be taken by the causer (in this case 'I') and since the direct object slot is already taken (in this case by 'the book'), Hasan moves to the next available slot, that of the indirect object.

An **analytic causative** is one where a separate verb is used to get the notion of causation, as in English *to make someone do something*.[157]

The difference between these three types of causatives is not merely in form, but also in concept. A lexical causative expresses a more direct causation, or a tighter link between cause and event, than an analytic causative. Compare the difference between the following sentences: *John killed the dog* versus *John caused the dog to die*. In the former sentence we may assume that John was directly responsible for the death of the dog. In the latter, however, John is further removed from the event of the dog's death and it might not necessarily be intentional. For instance, John might inadvertently have given the dog something to eat which it was highly allergic to. Or John might have used a fertilizer for his flowerbeds that happened to smell so incitingly that the dog ate some and was poisoned by it. In the former case we have an instance of **direct causation**, while in the latter we have an instance of **indirect causation**. Haiman (1983) proposed that the form of the expression and the directness correlate: the longer the morphological form of the construction, the more indirect the causation, or conversely, the shorter the morphological form of the causative construction, the more direct the causation. This would make lexical causatives express more direct causation while analytic causatives would express more indirect causation, and morphological causatives would come somewhere in between.

The above also implies varying degrees of control over the caused events. Languages may code the causee differently morphologically to differentiate between degrees of control of the causee. For instance in English the nominative case is used when the causee has a high degree of control, while the accusative is used when the causee has a lower degree of control. Compare the sentences *I asked that he leave* with *I asked him to leave*, where in the former the causee (*he*) has a right to say no (high degree of control), but in the latter the causee (*him*) is not likely to have the option of saying no (low degree of control). The causee has no control at all in a sentence like *I made him leave*. In Japanese this is expressed through different postpositions (case markers):

Japanese (Japanese (Japanese): Japan)

(186) a. Taroo ga ik-u
PN NOM go-PRES
'Taro goes.'

157. I am in fact simplifying matters greatly here by including analytic causatives as a valency increasing device. It is a matter of contention if a construction like the English *to make someone do something* belongs to the nonperiphrastic causative type (where two events are merged into one clause of a complex event) or to the periphrastic type (where the two events constitute two clauses). For more details on periphrastic causatives and how they differ from nonperiphrastic ones, see, for example, Song (1996, 2011a and 2011b) and the references therein.

b. Hanako ga Taroo o ika-se-ta
PN NOM PN ACC go-CAUS-PAST
'Hanako made Taro go.'
c. Hanako ga Taroo ni ika-se-ta
PN NOM PN AG go-CAUS-PAST
'Hanako had Taro go.' (Shibatani 1990: 308f)

In (186a) we have an intransitive clause with only one argument. In (186b and c) the clause is made transitive by way of the causative construction (the suffix *-se*). The difference between (186b and c) is that in (186b) Taro, marked in the accusative case, has minimal control (is not able to refuse to go), while in (186c) Taro, marked in the agentive case, has more control (Taro may choose not to go). In (186c) Taro was persuaded to go, while in (b) Taro was basically forced to go.

9.2.1.2 *Applicatives*

An **applicative** construction increases the valency of a base clause by one by making an optional argument (adjunct) into an obligatory argument (object) through a marker on the base clause verb. The new object of the clause is called the **applied object**. If the applicative is formed from a transitive base clause this means that the base clause already has an object. This "original" object is called the **basic object** in applicative constructions, to distinguish it from the new, applicative, object. For a very thorough discussion on applicative constructions, see Peterson (2007). Tukang Besi is an example of a language with applicative constructions:

Tukang Besi (Austronesian (Sulawesi): Indonesia)

(187) a. no-ala te kau
3.REAL-fetch the wood
'She fetched the wood.'
b. no-ala-ako te ina-su te kau
3.REAL-fetch-APPL the mother-my the wood
'She fetched the wood (as a favour) for my mother.' (Donohue 1999: 256)

In (187a) we have an ordinary transitive, a two-place predicate, demanding two arguments, the subject (indicated as a person affix *no-* 's/he' on the verb) and the direct object (*kau* 'wood'). With the applicative suffix *-ako* in (187) we get a clause containing a three-place predicate, i.e. a verb demanding three arguments, the subject and the direct object, both of which stay the same from the base clause, and the applied object (in this case a benefactive), *inasu* 'my.mother'. A sentence like **noalaako te kau* 'I fetch.APPLicative the wood' would not be acceptable, as the necessary argument of who I am fetching it for is missing.

In her sample of 183 languages Polinsky (2011b) found 83 (45.4%) languages with applicative constructions of varying kinds. Of these, the majority (74 languages or 40.4% of the entire sample) allow applicatives for both transitive and intransitive clauses, while seven languages (3.8% of the entire sample) restrict applicative constructions to transitive clauses, and only two languages (1.1% of the entire sample), Fijian and Wambaya (Australian (West Barkly): Australia), restrict applicative constructions to intransitive clauses. Furthermore, languages may differ in terms of which role gets promoted to an applicative object. In the Tukang Besi example above it was the benefactive. In Yagua, however, the instrumental may be promoted.

Yagua (Peba-Yaguan (Pea-Yaguan): Peru)

(188) a. sịịchtiñíí quiivą́ quichitya
sa-jịchitiy-níí quiivą́ quichiy-tya
3SG-poke-3SG fish knife-INSTR
'He pokes the fish with a/the knife.'

b. sịịchtiyara quichiy
sa-jịchitiy-ta-rà quichiy
3SG-poke-APPL-INAN.OBJ knife
'He pokes it with a/the knife.' (Payne 1985a: 272f)

In (188) the instrumental of sentence (a) becomes an obligatory argument of sentence (b) because of the applicative suffix *-ta*. Of the languages in Polinsky's sample that have applicative constructions, most (51 or 27.9% of the entire sample) allow applicative objects for arguments with a benefactive role as well as arguments with other semantic roles. A number of languages (36 or 19.7% of the entire sample) allow applicative objects to be formed with benefactive roles only. The common semantic roles for the applicative object are locatives (18 languages or 9.8% of the entire sample), instruments (17 languages or 9.3% of the entire sample) or both locatives and instruments (12 languages or 6.6% of the entire sample) (Polinsky 2011b). Applicatives tend to be found in areas with languages with rich verbal morphology, such as sub-Saharan Africa (especially the Bantu languages), Papua New Guinea and northern Australia, as well as North and Meso-America. The Eurasian continent almost completely lacks applicative constructions, except for a cluster in the Caucasus area, which also correlates with languages that have rich verbal morphology.

9.2.2 Reducing valency

Reducing the valency of a clause means employing devices that render an obligatory participant optional. Crucially, the omission of one of the arguments from the base clause in the valency reduced clause does not make the new (valency reduced) clause ungrammatical.

9.2.2.1 *Passives*

A **passive** construction reduces the number of arguments by demoting the subject of the base clause, the active counterpart clause. The subject of the active counterpart clause thus becomes either an optional participant (an adjunct) or is omitted entirely. The sentence *John hit the ball* is an example of an active transitive clause where *John* is the subject (or more exactly the A of the clause) and *the ball* is the object (or more exactly the P of the clause). This clause can be made intransitive by passivizing it to *The ball was hit* (*by John*), where the previously necessary argument *John* is now only an optional argument and may be omitted. If a ditransitive clause is passivized and loses one argument, it thus becomes a simple transitive clause (compare *I gave her a book* with *A book was given to her (by me)*).[158] Active and passive constructions belong to the category of grammatical **voice** (or **diathesis**), which, very simplified, basically announces the semantic role of the subject in the clause. In the active voice the subject is the agent or actor, while in the passive the subject is the patient, recipient or benefactive. For a thorough discussion on grammatical voice in general, see Klaiman (1991). For thorough and accessible discussions on the passive in particular, see, for example, Keenan & Dryer (2007) and the classic piece of work by Siewierska (1984).

Typical features of the passive are, as hinted above, that it has an active counterpart and that the subject of the active counterpart is rendered non-obligatory or even omitted completely in the passive counterpart clause. Another typical feature of the passive is that its subject corresponds to the direct object of the active counterpart. Yet another typical feature of the passive is that it tends to be used as a foregrounding device: in the sentences *John hit the ball* and *I gave her a book* the direct objects are not the topics of the clauses. With passivization the direct objects get topicalized: in *The ball was hit* and *A book was given to her* the direct objects of the active counterparts have become the topics of the passive clauses.

Passives may be expressed either morphologically (synthetically) or analytically (periphrastically). English is an example of a language with an analytic passive, where the active verb is passivized through the construction auxiliary + verb.participle, as in BE + VB-*en* (e.g. *was given*). Mualang is an example of a language with a morphological passive, where voice is marked by affixation on the verb.

---

158. The agent of the clause may thus optionally be either overt or covert: in *The ball was hit by John* and *The book was given to her by me* the agents (*John* and *me* respectively) are overt. In *The ball was hit* or *The book was given to her*, the agent is covert but is implied – even if the agent of the action is supressed there is still an implication that the ball was hit by someone and that the book was given by someone.

Mualang (Austronesian (Malayic): Indonesia)

(189) a. urang N-curi manuk ku
person ACT-steal chicken 1SG
'Somebody stole my chicken.'
b. manuk ku da-curi
chicken 1SG PASS-steal
'My chicken was stolen.' (Tjia 2007: 151)

In (189) the verb of the active clause in (a) has been passivized in (b) by the passive prefix *da-*. Notice also that the subject (*urang* 'somebody') in the active clause (a) has been demoted and omitted, while the object (*manuk* 'chicken') of (a) has been promoted to the subject of the passive clause in (b). The two arguments of the active clause have thus been reduced to one argument in the passive clause.

The kinds of passives discussed so far typically involve the demotion of the most agent-like argument and the promotion of the most patient-like argument of the clause. The resulting passive thus has an overt lexical subject (the promoted patient-like argument). These kinds of passives are called **personal passives**. In the kinds of passives that only involve the demotion of the agent-like argument, but not the promotion of the patient-like argument, the overt lexical subject is lacking. These kinds of passives are called **impersonal passives**. An example of an impersonal passive, where there is no lexical subject, can be found in German:

German (Indo-European (Germanic): Germany)

(190) Hier wird ge-tanz-t
here be PTCP-dance-PTCP
'Dancing takes place here.' (lit. 'here is danced') (source: personal knowledge)

In (190) there is no lexical subject, but only the passive form of the verb *tanzen* 'to dance'.

It is somewhat more common cross-linguistically for languages to lack passives than to have them. In her sample of 373 languages Siewierska (2011c) found that 162 languages (43.4%) have passives while 211 (56.6%) lack them. Passives are very common in Eurasia, Eastern Africa, Eastern North America and northern South America while they are relatively uncommon in coastal West Africa, Southeast Asia and northern Australia, Meso-America and eastern North America. The figures change insignificantly once the contact languages Ndyuka and Sango (both lacking passives) have been subtracted from Siewierska's sample. The pattern differs in APiCS, where only 23 languages (20.3%) lack the passive altogether (Michaelis et al. 2013: feature 90). It thus seems that pidgin and creole languages are more likely to have passives than non-creoles.

9.2.2.2 *Antipassives*

The term **antipassive** was coined by Silverstein (1972: 395) to capture the fact that the construction is basically a mirror image of the passive. Just as with passives the antipassive construction makes one of the two necessary arguments of the base clause optional. However, while in passives the agent-like argument of the active clause gets demoted and the patient-like argument of the clause gets promoted, in antipassives it is the patient-like argument that gets demoted while the agent-like argument stays. The crucial thing for both passives and antipassives is that valency is reduced. Thus a two-place verb like *throw*, which normally requires two arguments (the entity that does the throwing and the entity that gets thrown), only requires one argument if it gets passivized (as in *The ball was thrown*). Likewise, with antipassives, a two-place verb can be modified through antipassivization to require only one argument. The difference between the two constructions lies in which argument it is that stays and which argument it is that becomes optional or even suppressed. As with passives, the reason for using an antipassive construction very often lies in pragmatic factors of focus and topicalization, as one participant is downplayed and the other is highlighted. Consider the following example:

Chukchi (Chukotko-Kamchatkan (Northern Chukotko-Kamchatkan): Russia)

(191) a. ʔaaček-a kimiʔ-ən ne-nlʔetet-ən
youth-ERG load-ABS 3PL.SUBJ-carry-AOR.3SG.OBJ
'(The) young men carried away the/a load.'

b. ʔaaček-ət ine-nlʔetet-gʔə-t kimitʔ-e
youth-ABS ANTIP-carry-AOR.3SG.SUBJ-PL load-INSTR
'(The) young men carried away the/a load.' (Kozinsky et al. 1988: 652)

In (191a) we have a transitive clause and the two-place verb is marked for both arguments, with *ne-* indexing the A (*ʔaačeka* 'youth.ERGative') and the portmanteau *-ən* signalling the aorist as well as indexing the P (*kimiʔən* 'load.ABSolutive'). In (191b) the clause only has one argument, as the antipassive form of the verb shows by being indexed only for the single argument of the clause (*ʔaačekət* 'youth.ABSolutive') with the portmanteau suffix *-gʔə* and the plural suffix *-t*. The participant that was originally the patient-like argument has now become an adjunct, as is also signalled by the instrumental case marking (*-e*). Very often antipassives generalize the meaning of the patient-like argument of the base clause. Thus what in (191a) might refer to specific loads, may in (191b) have been "de-individualized" (they have become semantically generic) into loads in general. A possible way of trying to capture the difference in English could be to translate (191b) as 'The young men went load-carrying', or 'The young men engaged in load-carrying', where no specific loads are referred to.

Antipassives are much less common cross-linguistically than passives. In Polinsky's (2011a) sample of 194 languages, only 48 (24.7%) have antipassives while 146 (75.3%) lack them. Languages with antipassive constructions cluster mainly in the Caucasus, in northern Australia and in North America. It is often assumed that antipassives correlate with ergative alignment, but while there are more ergative languages than accusative ones with antipassives in Polinsky's sample there is "no principled correlation between ergativity and the antipassive" (2011a). Notice also that passives and antipassives are not mutually exclusive. For example, thirteen languages are listed in the WALS database as having both passive and antipassive constructions.[159]

9.2.2.3 *Noun incorporation*

Noun incorporation was discussed in 6.1.2. What it essentially means in terms of valency is that one of the arguments, typically the object, gets incorporated into the verb stem, thus creating a one-argument clause. An example of this can be found in Yucatec (52), repeated here for convenience:

Yucatec (Mayan (Mayan): Mexico)

a. t-in-p'oʔ-Ø-ah nòok
   COMPL-1SG-wash-it-PERF clothes
   'I washed (the) clothes.'

b. p'oʔ-nòok-n-ah-en
   wash-clothes-ANTIPASS-PFV-1SG.ABS
   'I clothes-washed.' (= 'I washed clothes.') (Bricker 1978: 15)

In (a) the noun *nòok* 'clothes' stands alone as an object to the verb *p'oʔ* 'wash'. Notice that the verb is not only inflected for the agent-like argument (*-in-* '1SG'), but also for the patient-like argument of the clause (*-Ø-* 'it'). In (b) the noun has been incorporated into the verb; notice that the verb now only carries an inflection for one argument (*-en* '1SG.absolutive'). The clause has thus been detransitivized and now only demands one argument. A possible way of trying to capture the decrease in valency is in the English translation 'I clothes-washed'.

159. These are Basque, Koyraboro Senni (Nilo-Saharan (Songhay): Mali), Krongo, Päri, Chamorro, Yukulta, Upriver Halkomelem (Salishan (Central Salish): Canada), Nez Perce, Comanche, West Greenlandic, Jakaltek (Mayan (Mayan): Guatemala), Tzutujil and Sanuma (Yanomam (Yanomam): Venezuela).

VALENCY INCREASING INCORPORATION

Nadëb (Nadahup (Nadahup): Brazil) allows postposition incorporation (see also Example (54) above) which transitivizes an intransitive verb, promoting the argument of the postposition to an object and making the original subject of the intransitive verb an agent. Postposition incorporation is also possible for transitive verbs in Makú, in which case again the argument of the postposition is promoted to an object, and the original object gets delegated to the periphery (Martins & Martins 1999: 262).

9.2.2.4 *Anticausatives*

Anticausatives, also called **middles** or **mediopassives**, are the logical opposites of causatives. With causative constructions we have a non-causative verb and add an element to make it causative, thereby increasing the valency by one. With anticausatives it is exactly the opposite: we start with a causative base and add an element to take out the sense of causation, thereby reducing the valency by one (the causer is no longer needed). The reason for the term 'middle' for anticausatives is because the constructions are neither passive nor active; in active clauses we have the agent doing something to the patient. In passive clauses we have something happening to the patient but the identity of the agent is downplayed. In middles, on the other hand, the role of the agent is completely ignored, and the situation is treated more like a process. With passives we have a covert agent: in *The ball was thrown* there is an implication that someone threw the ball. Thus a sentence like *The ball was deliberately thrown* is acceptable, since even if the identity of the agent is only optional or even suppressed, we may still assume that there is an agent involved. In other words, the clause is semantically transitive even if its valency has been reduced syntactically. With anticausatives (or middles), however, there is no implied agent for the situation. Compare the following sentences:

(192) a. The man broke the vase.
b. The vase broke.

In (192a) we have an active transitive clause; the passive equivalent is *The vase was broken (by the man)*. Sentences like *The man deliberately broke the vase* and *The vase was deliberately broken (by the man)* are quite acceptable. In (192b), however, we have an example of an anticausative (or middle), where no agent is implied – something like **The vase broke (by the man)* is not possible. Furthermore, the lack of the implied agent precludes any sense of volition: something like **The vase broke deliberately* is not acceptable.

English does not have any morphological marking for the anticausative (or middle) voice, but many languages do. Yagua is an example of such a language:

Yagua (Peba-Yaguan (Peba-Yaguan): Peru)

(193) a. sa-nóóta-máá-rà
3SG-knock.down-PERF-INAN
'S/he has knocked it down.'

b. sa-nóóta-y-máá
3SG-knock.down-ANTICAUS-PERF
'S/he has fallen down.' (Payne 1985a: 62f)

In (193a) we have an active transitive clause, as indicated by the verbal prefix *sa-* referring to the agent-like argument 's/he' as well as the verbal suffix *-rà* referring to the inanimate patient-like argument 'it'. In (193b), however, the anticausative suffix *-y* not only takes out one argument (there is no P anymore), but also takes out the sense of causation, which can only be captured in English by changing the verb in the translation from 'knock down' to 'fall down'. Notice that the Yagua base verb *nóóta* 'knock. down' has not been replaced by any new verb, but merely been modified by the anticausative suffix *-y*.

9.2.3 Transposing valency

With valency transposing devices the number of arguments of a clause is not changed, but the roles between them is somehow altered. This means that the verb still requires the same number of arguments as in the base clause to make the clause grammatical. It is simply that their respective semantic roles get swapped. As with valency reducing or increasing devices, the main motivation for these operations are such pragmatic factors as topic and focus, and so on.

9.2.3.1 *Inversion*

With **inversion** the alignment between the arguments in the base (or direct) clause has been exchanged (or inverted). Consider the two sentences from Nocte:

Nocte (Sino-Tibetan (Northern Naga): India)

(194) a. nga-ma ate hetho-ang
I-ERG he teach-1SG
'I will teach him.'

b. ate-ma nga-nang hetho-h-ang
he-ERG I-ACC teach-INV-1SG
'He will teach me.' (DeLancey 1981: 641 citing Das Gupta 1971)

The verbal ending agrees with whichever argument is higher on the animacy hierarchy, in this case the first person. The default is that whichever argument that is higher on the animacy hierarchy is interpreted as the subject of the clause, i.e. the

first person in (194a). With the inverse marker *-h* this is switched so that the subject is interpreted to be whichever argument is lower on the animacy hierarchy, in this case the third person. Notice that the verbal ending still agrees with the first person, which is now the P of the clause (*nganang* 'I.ACCusative'). This is also supported by the ergative case marking *-ma*, which appears on the first person in (194a) but on the third person in (b). The purpose of this is basically one of topicality; whichever argument is intended to be more topical will be placed in the subject (or, more accurately, the A) position.

Some languages, such as Plains Cree, mark both the direct and the inverse clause:

Plains Cree (Algic (Algonquian): Canada)

(195) a. ni-wa·pam-im-a·-w-a o-kosis-a
1-see-OBV-DIR-3-OBV 3-son-OBV
'I see his son.'

b. ni-wa·pam-iko-yi-w-a o-kosis-a
1-see-INV-OBV-3-OBV 3-son-OBV
'His son sees me.' (Dahlstrom 1986: 43)

In both sentences in Example (195) there are two arguments, one which is unmarked, the **proximate**, and one which is marked, the **obviative**. With the direct marking it is the proximate which functions as the most agent-like argument of the clause (also called the 'actor' in literature discussing inverse constructions and obviatives), in this case the first person. The argument marked in the obviative is thus the patient-like argument (also called the 'goal' in literature discussing inverse constructions and obviatives). With the inverse marking this relationship is swapped: the argument marked in the obviative (*okosisa* 'his son.obviative') is interpreted as the A (the 'actor') of the clause while the proximate is interpreted as the P (the 'goal'). The difference between the two sentences is not quite possible to capture in English; the closest equivalent would be something like *I see his son* (a) versus *I am seen by his son* (b).

9.2.3.2 *Dative shift*

Dative shift, a strategy found in English but which is not cross-linguistically common, seemingly realigns the two objects of a ditransitive clause. Compare the English sentences *I gave the book to him* with *I gave him the book*. In the first sentence we have a direct object (*the book*) and an indirect object (*to him*), flagged by the preposition *to*. In the second sentence we have what looks like two direct objects (*the book* and *him*), that is, the two objects formally look the same, called **double object**, as mentioned above (see 9.1.3.4). We saw above that about a fifth of the languages in Haspelmath's (2011a) sample have double objects, while only about ten percent have a mixed system where a ditransitive clause may have either one direct and one indirect object, or two

direct objects. Notice that dative shift typically involves the semantic roles of recipient or benefactive.

Dative shift does not alter the number of participants in the clause; the motivation for choosing one construction over the other again lies in pragmatic factors such as topicality.[160] However, even if there is a difference in topicality between *I gave the book to him* and *I gave him the book*, the number of obligatory participants remains the same and as such dative shift is a valency transposing device. Something like **I gave him* or ??*I gave the book* are not acceptable, except possibly the latter under specific elliptic contexts (such as an answer to the question *What did you end up giving him, the book or the film? – I gave the book*).

9.3 Simple clauses in sign languages

Alignment of clause participants in sign languages is typically neutral, that is, there is no overt marking differentiating between the S-argument of an intransitive clause from the A-argument or P-argument of a transitive clause. These arguments all look formally the same, irrespective of what grammatical relation or semantic role they have in the clause, as is the case with spoken English nouns.

Due to the nature of sign languages, verbs (or event signs) show a much greater degree of modification in accordance with the participants of the clause than spoken languages do. Depending on which participant is the subject (or A) of the clause and which is the object (or P) of the clause, the spatial movement of the verb will be modified accordingly. Thus a sentence like *I help you* will be formed by signing the movement towards the object (or P) of the clause, *you*. The movement of the sign will be reversed if the sentence is *You help me*, with the movement going from the A (*you*) and towards the P (*me*).

A very common feature with sign languages, in fact probably universal, is the use of verbal classifiers, as mentioned in 5.3. Here the verb (or event sign) is modified to incorporate the object of the clause. In other words, if the basic sign for GIVE is both palms turned upwards + movement from the body outwards, a sentence like *I give you a book* will be formed by modifying the event sign GIVE to have the hand shape resemble the sign for book, while the movement of give will remain the same.

---

160. This is again a simplification. It could also be argued that dative shift may function as a valency increasing device, where a bivalent verb such as *write* is made trivalent: compare *I wrote a letter* (two arguments, *I* and *letter*) with *I wrote her a letter* (three arguments: *I*, *her* and *letter*). Seen this way the dative shift construction differs from applicative constructions (discussed in Section 9.2.1.2) only in that the valency increasing strategy involves constituent order and case-marking with dative shift, while with applicatives it also involves verbal marking.

The sentence *I give you a book* will thus be expressed something like I BOOK-GIVE YOU, where the event sign BOOK-GIVE signifies 'to give a book'. The spatial movement of this complex sign may then be altered according to the syntactic roles of the participants. In other words, the motion of the event verb will begin at the location of the subject (A) participant of the clause and end with the indirect object participant, or recipient (R). This could be analysed as an instance of object agreement, i.e. that the verb shows agreement with the patient-like argument of the clause as opposed to the agent-like argument of the clause, a very rare strategy among spoken languages. The event signs that are possible to modify in this way are often called 'agreement verbs' (see, for example, Meir 2002). With event signs that cannot be modified with classifiers ('plain verbs') the grammatical relations of the clause are usually, but not always, determined through word order.

Some sign languages avoid having more than one nominal participant in the clause, instead chopping up the situations described into several intransitive propositions. IPSL is an example of such a language, where a sentence like *The man lied to his wife* would be expressed as *The man lied. His wife was angry*. However, transitive clauses where one participant is expressed with a pronominal are frequent (Zeshan 2003c: 170). A similar preference for one-argument clauses can be found in, for example, ISL and ABSL (Meir 2010).

An example of a valency increasing causative can be found in Greek Sign Language (ENG), where the verb GIVE may be used as an auxiliary to express causation:

ENG (Sign Language: Greece)

(196) $\text{INDEX}_2\ {}_2\text{GIVE-AUX}_3$ BURDEN END
'Stop being a trouble/nuisance to him/her!' (Pfau & Steinbach 2006a: 46)

The auxiliary is used only with intransitives and "transitive psych-verbs", i.e. verbs that describe being in a specific psychological state (Pfau & Steinbach 2006a: 46). In Example (196) the psych-verb BURDEN thus means that 'X is in mental a state of discomfort'. The auxiliary GIVE[161] causativizes the clause and adds the meaning that 'someone causes X to be in a mental state of discomfort'. An obligatory participant (the causer) has thus been added to the clause.

While valency reducing devices such as passives are not generally reported for sign languages, Janzen et al. (2001) argue that the object agreement mentioned above actually is an instance of passive formation in ASL. Future macro surveys on possible valency alternating devices in sign languages will shed more light on the fascinating topic whether the mode of communication affects the types of clauses uttered.

161. The difference between the full verb GIVE and the auxiliary GIVE lies mainly in the lack of non-manual markers for the latter.

9.4 Summary

The elements that make up a simple clause have various roles and functions. Participants have different semantic roles in a clause, depending on how they develop the meaning of the clause. Participants also have different pragmatic roles, reflecting the information flow in the clause. Finally, participants have different syntactic roles, depending on which function they have in the grammar of the clause.

In an intransitive clause there is only one argument, the subject (S), while in a transitive clause there are two arguments, the most agent-like argument (A) and the most patient-like argument (P). Some languages also have avalent clauses, with no arguments. Languages have different ways of aligning S, A and P. Languages also differ in how they align the recipient-like argument (R) and the theme-like argument (T) in ditransitive clauses with the P of monotransitive clauses. Some languages have split systems, where alignments differ in different contexts. Furthermore, languages differ as to which arguments, if any, they index on the verb.

The valency of the verb can be altered in various ways. Causative constructions increase the valency by adding a concept of causation, making the argument of causer necessary. Applicatives increase the valency by turning an adjunct into an argument. Passives and antipassives reduce the valency by one, the former by making the A argument optional and the latter by making the P argument optional. Noun incorporation reduces the valency by incorporating one of the arguments into the verbal stem, while anticausatives reduce the valency by taking out the concept of causation. Inversion transposes the valency by swapping the syntactic roles of the A and P arguments, while dative shift transposes the valency by turning a ditransitive clause into a double object clause.

Sign languages generally have a neutral alignment. They very commonly, or even universally, have a subset of verbs that have object agreement in the form of verbal classifiers. Sign languages frequently show a preference for single argument clauses, except when one or both of the arguments is a pronoun. While valency adjusting operations do not seem to play a prominent role in sign language grammars, causative constructions have been reported.

9.5 Keywords

alignment
anticausative
antipassive
applicative
argument
causative
dative shift
inversion
participant roles
passive
pragmatic roles
predicate
semantic roles
split systems
syntactic roles
transitivity
verb agreement

9.6 Exercises

1. Identify the alignment of the languages below. How do pidgin and creole languages compare with non-creoles with respect to argument alignment?

   Swedish (Indo-European (Germanic): Sweden)

   (a) jag sov
   1SG slept
   'I slept.'

   han bar mig
   3SG carried 1SG
   'He carried me.'

   jag bar hit sakerna
   1SG carried here thing.PL.DEF
   'I carried the things here.' (source: personal knowledge)

   Araona (Tacanan (Tacanan): Bolivia)

   (b) ema táhui
   1SG slept
   'I slept.'

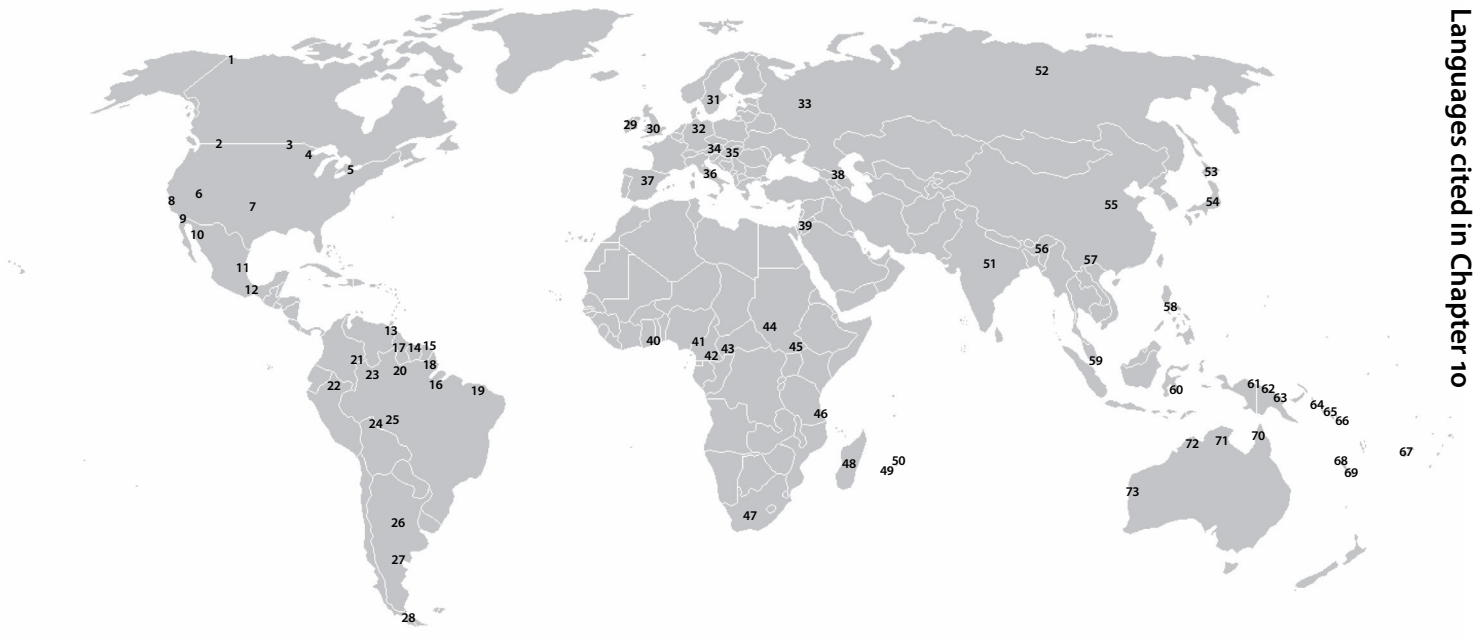
   huada ema dobeataiqui jidyo
   3SG 1SG carried here
   'He carried me here.'

   yama aicana jidyo dóbea
   1SG thing.PL here carried
   'I carried the things here.' (Pitman 1980: 15, 81, 83)

2. What are **split intransitive** and **split ergative** systems? How might **isomorphism** explain possible motivations for splits?
3. How do spoken and signed languages compare with respect to verb agreement?
4. What is the difference between **causative** and **anticausative**?
5. Is the following statement true or false? Motivate your answer.

   Alignment of clause participants can be adequately described in terms of grammatical relations alone.

Languages cited in Chapter 10



1 Eskimo Pidgin
2 Kutenai
3 Michif
4 Ojibwe, Minnesota
5 Cayuga
6 Navajo
7 Choctaw
8 Chumash, Barbareño
Salinan
9 Diegueño, Mesa Grande
Tiipay, Jamul
10 Seri
11 Tepehua
12 Zapotec, San Bartholomé Zoogocho
13 Warao
14 Ndyuka
15 Guyanais
16 Asuriní
17 Macushi
18 Tiriyo
Wayampi
19 Urubú Sign Language
20 Hixkaryana
21 Cubeo
22 Urarina
23 Nadëb
24 Wari'
25 Karó, Arára
26 Argentine Sign Language
27 Gününa Küne
28 Selknam
29 Irish
30 British Sign Language
English
31 Swedish Sign Language
32 German
German Sign Language
33 Russian
34 Austrian Sign Language
35 Hungarian
36 Italian
37 Spanish
38 Khwarshi
39 Jordanian Sign Language
40 Adamorobe Sign Language
41 Mbili
42 Konzime
43 Sango
44 Dinka, Agar
45 Päri
46 Matuumbi
47 Afrikaans
48 Malagasy
49 Réunion Creole
50 Mauritian Creole
51 Indo-Pakistani Sign Language
52 Evenki
53 Ainu
54 Japanese Sign Language
55 Chinese Sign Language
56 Assamese
57 Yi
58 Chabacano, Caviteño
Chabacano, Ternateño
Chabacano, Zamboanga
59 Singlish
60 Bajau, Sama
61 Tobati
62 Yimas-Arafundi Pidgin
63 Mugil
64 Roviana
65 Gela
66 Buma
67 Fijian
Pidgin Hindustani
68 Cèmuhî
69 Tayo
70 Wik Ngathana
71 Mangarrayi
72 Ungarinjin
73 Nhanda

Chapter 10

Word order

We have seen that sentences consist of units such as verbs and their arguments. Languages differ in the way these units, or constituents, are arranged relative to one another, as well as to what extent this arrangement is fixed or flexible. Syntax deals with how meaningful units may be ordered, both on the clausal and on the phrasal level. This chapter gives a very brief introductory note on some key concepts in syntax (10.1) before discussing the typology of word order, that is, the order of the meaningful elements in a clause or phrase (10.2.1 and 10.2.2). Examples of some typical word order correlations are given (10.2.3), as well as examples of languages where the word order is determined by other factors than grammatical relations (10.2.4). Section 10.3 gives an overview of issues relating to word order in sign languages.

10.1 A very brief introductory note on syntax

We have seen that smaller units, phonemes and morphemes, may combine to form larger units, words, which in turn may combine to form even larger units, such as phrases, clauses and sentences. However, languages do not allow units to be thrown together haphazardly. For example, noun phrases like *the dog* or *the happy dog* are grammatical, but **dog the* and **dog the happy* are not, even though they as such contain the same information (it is a specific dog and it is happy). Likewise, while *The dog was chasing the cat* is a grammatical sentence in English, something like **Was chasing the cat the dog* is not. **Syntax**, which derives from the Greek word *sýntaxis* 'arrangement', deals with how words or a set of words are put together, or arranged, to form grammatical phrases, clauses and sentences. Specifically, syntax deals with the order or arrangement of constituents, which are the functional elements of any larger construction. This section is intended as a very brief introductory note on syntax, with the main purpose of presenting some essential terminology. It makes no claims whatsoever of presenting a comprehensive discussion of the various issues and traditions in studies of syntax. A very accessible introduction to syntax for the complete beginner is Tallerman (2005). For an extraordinarily detailed introduction to syntax, where a multitude of perspectives are included, see Givón (2001a and 2001b).

10.1.1 The building blocks of sentences

A sentence consists minimally of a clause, which in turn consists of phrases, which are made up of one or several words. As Huddleston & Pullum put it

> i. Sentences have parts, which may themselves have parts
> ii. The parts of sentences belong to a limited range of types
> iii. The parts have specific roles or functions within the larger parts they belong to
>
> (2002: 20)

The way languages tend to structure their lexicon into word classes, or parts-of-speech, was discussed in Chapter 6. Different types of parts-of-speech may take different syntactic functions in the next larger unit, the phrase.

10.1.1.1 *Phrases*

A phrase minimally consists of one word, but may also be a group of words that belong together to form one syntactic unit. The entire phrase is classified according to the category that the head of the phrase belongs to and usually takes the syntactic position that the head would take. Thus noun phrases, for example, where the head is a noun (or a substitute for a noun, such as a pronoun), take the positions and functions that a noun has in the clause (see Chapter 7). Likewise, verb phrases, where the head is a verb, take the positions and functions that a verb has in the clause (see Chapter 8). It is very important here to keep in mind that while a constituent may minimally consist of one word, it may also consist of several words or even whole clauses. Thus the noun phrase *the dog* consists of two constituents, *the* and *dog*. This noun phrase may in turn be a constituent in the clause *The dog was barking*, which also consists of two constituents (*the dog* and *was barking*). Further, the entire clause may be a constituent, as in the sentence *The man heard that the dog was barking*. Languages may have other kinds of phrases than NPs and VPs. What follows are some of the most commonly distinguished ones.

Adjective phrases (AP or AdjP) have an adjective as its head, for example *young* in *The young doctor was just about to treat his first patient*. Adjective phrases may contain modifiers to the head, such as *very* in *very young* (as in, for example, *The very young doctor was just about to treat his first patient*). An example of an entire clause that functions as a modifier to an adjective phrase is *that his diagnosis will be wrong* in *The doctor is afraid that his diagnosis will be wrong* where the underlined sequence constitutes an AdjP.

Adverb phrases (AdvP) have an adverb as a head, for example, *quickly* in *The man ran quickly* and may also contain modifiers, such as *very* in *The man ran very quickly* (where the AdvP is underlined).

Adpositional phrases have an adposition as its head, most commonly either a preposition or a postposition (either PP for both types or, in order to distinguish between the two, PrepP versus PostP), such as *after* in *The man ran after the bus*. Adpositional phrases may also consist of several words, for example *until after* in *You'll have to wait until after dinner*, or entire clauses, such as *until they've finished their game* in *You'll have to wait until they've finished their game* (where the PP is underlined).

10.1.1.2 *Clauses*

As mentioned in Chapter 9, a clause will consist minimally of a predicate and the arguments of that predicate. This means that if the verb requires one or more argument(s), the clause will not be grammatical without these arguments.

A simple sentence consists only of one **main clause**. *The man saw the dog* is an example of a simple sentence. A complex sentence consists either of two or more coordinated main clauses, or a main clause and one or more **subordinate clauses**. A main clause is, very simplified, a clause that can function on its own, i.e. which is independent of any other clause. A subordinate clause, on the other hand, cannot stand on its own, but is dependent on the main clause. Examples of subordinate clauses (underlined) are *The baby was crying because it was hungry* (adverbial clause), *The mother knew that the baby was hungry* (complement clause), *The nurse, who was very kind, helped the old man up the stairs* (relative clause). For more on complex sentences and various types of subordinate clauses, see Chapter 11.

10.1.2 Constituents

The above-mentioned building blocks form functional components, **constituents**, of larger structures. As mentioned, phrases generally take the syntactic function and position in the clause of their head. On the clause level, in addition to the core constituents already discussed, namely the verb and its arguments, a sentence may also have a complement and an adverbial. A **complement** is typically an attribute to an argument, such as *a doctor* in *The woman is a doctor*.[162] An **adverbial** is a constituent that functions as an adverb. While it may consist of an adverb, it does not necessarily have to. Examples of adverbials (underlined) are:

---

162. It may, however, also be defined as a subcategorized obligatory unit of a head, such as a direct object of a transitive verb. According to this definition, *a book* in *The man took a book* would be a complement to *took*, since *take* is a transitive verb. The reason that subjects are not generally considered complements according to this definition, even though they too are obligatory to a transitive verb, is because the subject in this definition is considered a clause-level constituent on par with the other clause-level constituent, the predicate.

We arrived early.	(the adverb *early*)
We arrived in the evening.	(the PrepP *in the evening*)
We arrived the following day.	(the NP *the following day*)
We arrived after the others had left.	(the subordinate clause *after the others had left*)

On the phrase level constituents typically function as modifiers. Examples of modifiers in the NP are adjectives (or AdjPs), such as *red* in *the red book*, numerals, such as *three* in *the three books*, adpositional phrases, such as *on the shelf* in *the book on the shelf*, and so on. Examples of modifiers in the VP are auxiliaries, such as *will* in *will run*, adverbs (or adverbials), such as *quickly* in *run quickly*, and so on. Examples of modifiers in the AdjP are adpositional phrases, such as *with the result* in *happy with the result*, or adverbs (or adverbials), such as *very* in *very happy*. Examples of modifiers in the AdvP are adverbs (or adverbials), such as *very* in *very quickly* or *somewhat* in *somewhat hastily*.

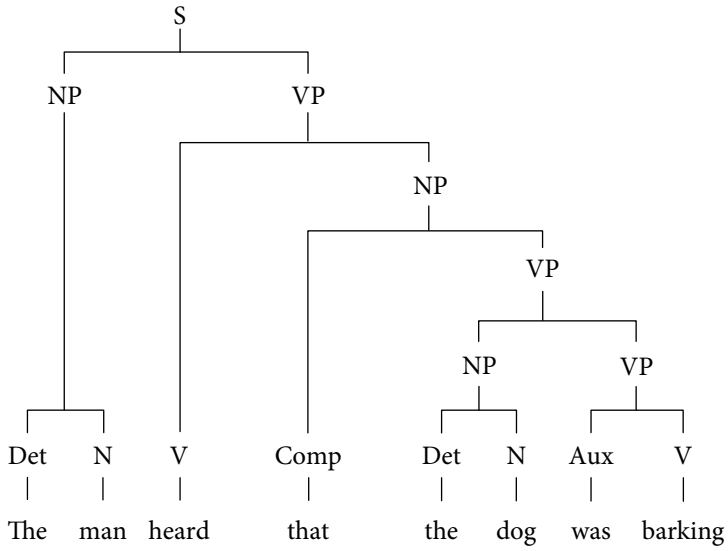
In order to keep track of how elements group together to form constituents, the internal structure of a phrase, clause or sentence may be illustrated through bracketing or tree diagrams. The sentence above, for instance, may be analysed as follows:

[[[The] [man]] [heard [that [[[the] [dog]] [[was] [barking]]]]]]

Each unit receives its own brackets until the constituents are not further divisible. The largest unit is the sentence [The man heard that the dog was barking], which consists of two units, the NP [The man] and the VP [heard that the dog was barking]. These units further consist of two units each: the definite article [the] and the noun [man] make up one unit, while the verb [heard] and the subordinate clause [that the dog was barking] make up the other. The latter unit, *that the dog was barking*, can be even further divided into the complementizer [that], the NP [the dog] and the VP [was barking]. Finally, [the dog] may be divided into the definite article [the] and the noun [dog], while [was barking] may be divided into the auxiliary [was] and the main verb [barking].

This analysis may also be rendered as a tree diagram, which gives the same information as the bracketing, but simply illustrates in it a different manner. The "tree" is upside-down, with the root at the top and the branches spreading downward. Each level is connected through a **branch** and each connecting point of two (or more) branches forms a **node**.

This kind of analysis, where each constituent is divided into exactly two parts until no further division is possible, starts with the general assumption that the two core constituents of the sentence at the highest level are a subject and a predicate. Notice that each node forms a constituent (which may or may not be further divisible into



smaller constituents); at no point are constituents divided and lumped together across constituent boundaries. For example, there is no single node with the sequence *man heard* or *that the* or *dog was*.[163]

10.2 Word order typology

Since the seminal study by Greenberg (1963, reprinted in Greenberg 1990) on the order of meaningful elements in languages, word order has been a highly prominent area of research in typology. Strictly speaking 'word order' actually refers to constituent order, and concerns constituents on both clausal and phrasal levels. For accessible overviews on word order typology, including the history of research on word order typology, see Song (2009) and (2010). See also Givón (2001a: 233ff).

Languages differ in how fixed, or rigid, their word order is. English is an example of a language with a rigid word order. For instance a sentence like *The child stole my money* is not normally possible to express as, for example, **My money stole the child*, **Stole the child my money* or **Stole my money the child*. This is because the syntactic roles of the constituents in English are determined by the word order: the subject

163. There are other ways of illustrating how the constituents relate to each other, depending on what tradition is followed. See, for example, Culicover & Jackendoff (2005), for a less hierarchical analysis, illustrated through so-called flat constituent structures.

comes before the verb and the object comes after it. Thus in the sentence *The dog chased the cat*, we assume that *the dog* is the subject of the sentence and *the cat* the object. If we swap the two NPs, giving the sentence *The cat chased the dog*, the two NPs by default also swap grammatical relations: *the cat* is the subject and *the dog* the object. However, not all languages are as rigid in their word order as English. Nhanda is an example of a language with free, or flexible, word (constituent) order.

Nhanda (Australian (Pama-Nyungan): Australia)

(197)

abarla-lu	wumba-yi	wur'a-tha
child-ERG	steal-PPERF	money-1SGOBL
S	V	O
abarla-lu	wur'a-tha	wumba-yi
S	O	V
wumba-yi	wur'a-tha	abarla-lu
V	O	S
wumba-yi	abarla-lu	wur'a-tha
V	S	O
wur'a-tha	wumba-yi	abarla-lu
O	V	S
wur'a-tha	abarla-lu	wumba-yi
O	S	V

'The child stole my money.' (Blevins 2001a: 125)

In (197) the case marking makes clear what syntactic roles the arguments have, so the meaning of the sentence stays the same even if the constituents are swapped around.

Having said this, it is worth to keep in mind that even languages with a rigid word order can be claimed to allow varied word orders. Consider the following:

(198)

a.	The child stole my money.	[SVO]
b.	The child my money stole.	[SOV]
c.	Stole it, the child did, my money.	[VSO]
d.	(There it was), stealing my money, the child.	[VOS]
e.	My money, the child stole.	[OSV]

Intuitively, any fluent speaker of English would consider (198a) an example of an ordinary, declarative, unmarked sentence. All the other sentences are somehow special. A sentence like (198b) would only be acceptable in a rhyming context. Sentences (198c–e) are rewritten in various ways, even the last one, where the comma marks a pause. In fact the last sentence is only acceptable in a contrastive context (for example, *It left the jewellery and silver, but my money, the child stole*.)

Word order typology seeks to map the basic constituent order in languages, which is not as straightforward to determine as one might initially think. First of all, in order

to determine the basic word order of a language, simple declarative sentences are sought, where both arguments of the verbs are nominal (i.e. consist of a noun and whatever constituents that belong to it, for example *the dog*, *the big dog*, and so on) and not pronominal. This is because pronominal arguments may follow different word order rules from nominal arguments. In Italian, for example, the pronoun may pre-cliticize to the verb, changing the word order from SVO to SOV.

Italian (Indo-European (Romance): Italy)

(199) a. il ragazzo ha visto la donna [SVO]
ART boy AUX see.PTCPL ART woman
S V O
'The boy has seen the woman.'

b. il ragazzo l'=ha visto [SOV]
ART boy ART=AUX see.PTCPL
S O V
'The boy has seen her.' (source: personal knowledge)

Apart from mapping sentences where the arguments of the verbs are nominal, the sentences also have to be simple declaratives, such as *Mary throws the ball* (for more on declaratives, see 12.1.1). As we saw in Example (198), pragmatically marked sentences, where, for example, items are rearranged due to such things as focus and topicality, may take a different word order from the most unmarked kind of declarative. Therefore it is the pragmatically neutral sentence that is considered to exhibit the basic word order. Another more marked sentence type is that of questions. In English, for example, the question word is sentence initial, irrespective of whether it refers to the subject or the object.[164] Compare *The man saw the ball* (SVO) with the questions *Who saw the ball?* (SVO) and *What did the man see*? (OSV). Again, the less marked sentence type, the declarative statement, is considered to exhibit the more basic word order. Often, though by no means always, this is also the most frequent word order in the language. Frequency is, in fact, the most straightforward factor in determining basic constituent order. This demands a large amount of varied texts. While it may be the case that the choice of text will affect the frequency of word order – different sets of texts might exhibit entirely different word orders – frequency is still a rather neutral operational test. In languages where one order is considerably more frequent than the others, anyone examining a large amount of texts is likely to arrive at the same conclusion. In most languages there is likely to be a dominant word order, but, as we will see, there are also languages where two or more word orders occur with roughly

164. Unless a preposition is involved, in which case the object is clause initial, as in *To whom did you give the ball*?

the same frequency. For more on how to identify the basic word order in a language, see Dryer (2007b: 73ff).

The following sections will discuss patterns of the dominant word order on both the clause level and phrase level. I am here following Dryer in using the expression 'dominant word order ', and not 'basic word order' so as "to emphasize that priority is given […] to the criterion of what is more frequent in language use, as reflected in texts" (2011b). For details on how Dryer determines the dominant word orders, see Dryer (2011b).

10.2.1 Order of clause constituents

There are six logical possibilities to order the core constituents subject (S), verb (V) and object (O) of a transitive clause: SVO, SOV, VSO, VOS, OSV and OVS. To make the discussion more comparable to other studies, I will refer to core clause constituents as 'subject', 'verb' and 'object', although these terms are used in a loose semantic sense where the 'subject' of a transitive clause denotes the most agent-like argument while the 'object' of a transitive clause denotes the most patient-like argument of the clause. As the discussion in Chapter 9 shows, a stricter and more correct notation would in fact be to list the six possibilities as AVP, APV, VAP, VPA, PAV and PVA.

All six word orders are attested in Dryer's (2011r) huge database of 1377 languages. A seventh group of languages consists of those where there is no dominant word order in terms of frequency in pragmatically neutral contexts. Table 10.1 summarizes the figures (ND stands for 'no dominant word order'):

Table 10.1 Word order patterns for three constituents: S, O and V (Dryer 2011r).

	N	%
SOV	565	41
SVO	488	35.4
VSO	95	6.9
VOS	25	1.8
OVS	11	0.8
OSV	4	0.3
ND	189	13.7
Total	1377	99.9

The vast majority of languages fall into the first two groups, where the subject is the first constituent of the three. Verb initial languages, the third and fourth groups, are rather rare. Object initial languages, the fifth and sixth group, are attested, but are very rare indeed. The group of languages listed as having no dominant order is actually

bigger than the verb and object initial languages taken together. Examples of the six different word orders, in descending order of size, are:

Ainu (Isolate: Japan)

(200) kamuy aynu rayke
bear person kill
S O V
'The bear killed the person.' (Shibatani 1990: 23)

Matuumbi (Niger-Congo (Bantoid): Tanzania)

(201) abụnwaásị aachéngịte ñụ́ụ́mba
PN he.built house
S V O
'Abumwas built a house.' (Odden 1996: 74)

Irish (Indo-European (Celtic): Ireland)

(202) tógann Máire an cat
lift.PRES PN ART cat
V S O
'Mary lifts the cat.' (Ó Dochartaigh 1992: 39)

Cèmuhî (Austronesian (Oceanic): New Caledonia)

(203) [ɛ ālī-hĩ] [ā-li mwà] [ɔ pā-li āpūlīp]
[3SG see-TR] [ART:NEUT-DEF house] [SUBJ ART:NF-DEF man]
V O S
'The man saw the house.' (Lynch 2002: 761)

Päri (Nilo-Saharan (Nilotic): Sudan)

(204) dháagɔ́ á-yàaŋ ùbúrr-ì
woman COMPL-insult PN-ERG
O V S
'Ubur insulted the woman'. (Andersen 1988: 292)

Warao (Isolate: Venezuela)

(205) erike hube abun-ae
PN snake bite-PAST
O S V
'A snake bit Enrique.' (Romero-Figueroa 1985: 107)

SOV word order is spread over the globe in Dryer's (2011r) sample, but is especially prominent in Asia (except Southeast Asia and the Middle East) as well as Papua New Guinea and Northern America (except the Pacific coast). SVO word order is

especially common in Europe, sub-Saharan Africa and Southeast Asia. Verb initial languages (VSO and VOS) are scattered around the world, but are very rare on the Eurasian continent (except for the Celtic languages in the far west of Europe). VOS order is not found at all on the Eurasian of African mainland in Dryer's sample. The object initial languages in the sample are not found on the Eurasian continent or in North America. Of the nine OVS languages, six are found in South America, Asuriní (Tupian (Tupi-Guaraní)), Hixkaryana and Tiriyo in Brazil, Cubeo (Tucanoan (Tucanoan)) in Colombia, Urarina (Isolate) in Peru and Selknam in Argentina, two in Australia (Mangarrayi (Australian: Mangarrayi) and Ungarinjin) and one in Sudan (Päri). Of the four OSV languages, two are found in South America (Nadëb in Brazil and Warao in Venezuela), one in Indonesia, Tobati (Austronesian (Oceanic)) and one in Australia, Wik Ngathana (Australian (Pama-Nyungan)).

The pattern for the APiCS languages differs considerably from that in WALS (where the proportions do not change after the two contact languages Ndyuka and Sango – both with SVO word order – have been subtracted from the sample). While all logical possibilities are attested (Michaelis et al. 2013: feature 1) , the vast majority of the languages, 62 of 76 (or 81.6%) have SVO as their dominant word order, while only six (7.9%) have SOV. Of the other logical possibilities, three languages (or 3.9%), Ternateño, Zamboanga and Caviteño, have VSO as their dominant word order and one (1.3%), Yimas-Arafundi Pidgin, has OSV as its dominant word order. Afrikaans (Indo-European (Germanic): South Africa), Michif, Pidgin Hindustani and Singlish (Creole (English-lexified): Singapore) have no dominant word order. VOS and OVS word orders are attested, but not as the dominant word order for any of the languages in the database.

Mapping languages for the constituents subject, object and verb (where, I repeat, the notions of subject and object are in a loose semantic sense of the most agent-like and most patient-like arguments respectively) requires enough examples of transitive sentences where both NP arguments are nominal. In fact, it is very often the case in available descriptions that one or both of the NP arguments in a transitive clause is a pronoun.[165] We have already seen (Example (199)) that languages may have a different word order for sentences where both arguments are nominal from sentences where one of the arguments is a pronoun. Very often this is due to the fact that the pronoun either cliticizes to the verb or is expressed exclusively through verb morphology. It may also be that the transitive clauses in available texts contain one covert argument,

165. Whether or not that reflects actual usage would be an interesting topic for further investigation. A large-scale corpus survey might show that, depending on the genre of the text, simple transitive declarative sentences more often contain at least one pronominal argument, or one covert argument, than that both arguments are full nouns.

that is, one argument that is not spelled out, but is inferred from the context (for instance via verb morphology). This often happens in Salinan, for example:

Salinan (Salinan (Salinan): USA)

(206) ra'm-hala' umck'a^{u}'yu
then-used talons
'Then (he) used his talons.' (Dryer 1989b: 42)

In (206) the subject is covert and is simply inferred from the context (such as the previous sentences in the narrative). A sentence like this, despite the fact that it has a transitive verb, does not give adequate information about the order of subject, object and verb.

Apart from the fact that transitive clauses where both arguments are nominal may not be easily available, mapping the word order of subject, object and verb also completely ignores intransitive clauses. One way of getting around both of these problems is to separately map languages for the order of subject and verb (which would then include intransitive clauses) and mapping the order of object and verb (which would make available all those instances where the subject is covert).

Of the 1498 languages in Dryer's (2011q) sample, the vast majority, 1194 languages (79.7%) have the dominant order subject-verb (SV) while 194 (13%) have the dominant order VS and 110 languages (7.3%) do not have any dominant order. Again the languages have been mapped for clauses where the subject is a nominal. The languages with VS are found largely in the same areas as VSO and VOS languages (cf. above). Examples of SV and VS orders follow:

Mbili (Niger-Congo (Bantoid): Cameroon)

(207) [a-bugu ya] a-yaŋa
[CP-hand my] CE-hurt
S V
'My hand hurts.' (Ayuninjam 1998: 327)

Roviana (Austronesian (Oceanic): Solomon Islands)

(208) taloa [se Zima]
leave [ABS PN]
V S
'Zima left.' (Corston-Oliver 2002: 491)

The majority of the languages in Dryer's (2011r) and (2011q) databases overlap in the sense that SOV, SVO and OSV languages are also SV languages, while VSO, VOS and OVS are also VS languages. However, with some languages the position of the subject relative to the verb differs between transitive and intransitive clauses. In these cases Dryer (2011q) lists the languages according to the order of their intransitive clauses.

In Tepehua, for example, transitive clauses where both arguments are nominals typically have SVO order, while in intransitives the nominal argument typically follows the verb (VS).

Tepehua (Totonacan (Totonacan): Mexico)

(209) a. ta-min-ta [pu:ma:-luw lapana:k-ni]
3PL.S-come-PFCT [CLASS-many people-PL]
V S
'A lot of people came.'

b. maqtali:-n ta-'u-y piyu:
wild.animal-PL 3PL.S-eat-IPFV chicken
S V O
'Wild animals eat chickens.' (Watters 1988: 460)

Another instance where Dryer's two maps do not match is when a language has been listed as having no dominant order for the constituents subject, object and verb, but where the reason for this is due to the fact that the position of the object differs relative to the verb. In other words, a language that has been listed as not having any dominant order because SVO and SOV are more or less equally frequent, will in the map dealing with only subject and verb be listed as having SV order. German is an example of such a language, due to the fact that the word order in subordinate clauses differs from that of main clauses.

German (Indo-European (Germanic): Germany)

(210) a. [der Mann] kauft [einen Ball]
[ART.M.NOM man] buy.3SG.PRES [ART.N.ACC ball]
S V O
'The man is buying a ball.'

b. ich glaube,
1SG.NOM think.1SG.PRES
dass [der Mann] [einen Ball] kauft
COMP [ART.M.NOM man] [ART.N.ACC book] buy.3SG.PRES
S O V
'I think that the man is buying a ball.' (source: personal knowledge)

In (210a) we have a main clause declarative sentence, with the typical SVO word order. However, if the clause is made into a subordinate, as in (210b), the verb is moved to sentence final position, which automatically changes the word order to SOV.

Dryer (20110) surveyed 1519 languages for their word order of the object and the verb. I repeat that the notion of 'object' here is in a loose semantic sense and that is would more accurately be described as the order of P and V. The languages are mapped for nominal (and not pronominal) objects, for the reasons discussed above.

However, for this map the subject may be pronominal or covert, since what is crucial here is simply the order between the verb and the (nominal) object. This fact alone makes more data available. The two logical orders OV and VO are found to be dominant in almost exactly the same amount of languages (713 or 46.9% with OV and 705 or 46.4% with VO word order), while 101 languages (6.6%) do not have any dominant order. OV languages are found in most of Asia, Papua New Guinea, North America and northern South America, as well as in a belt across Africa just under the Sahara. OV is also quite common in Australia. VO languages are predominantly found in Europe, Sub-Saharan Africa and Southeast Asia, Meso-America and along the Pacific North American coast. Examples of the two kinds of word orders follow:

Wayampi (Tupian (Tupi-Guaraní): Brazil)

(211) t-ata a-mɔɛnɨ́
ART-fire 1SG-light
O V
'I light the fire.' (Grenand 1980: 75)

Sama Bajau (Austronesian (Sama Bajau): Indonesia)

(212) kita-ku uggo'
see-1SG pig
V O
'I saw the pig.' (Donohue 1996: 784)

Again the languages in Dryer (2011r) and (20110) mostly overlap, in the sense that languages with SOV, OVS and OSV word order also have OV order and that languages with SVO, VSO and VOS order also have VO order. However, there are some mismatches between the two samples. First of all, there are some languages that are listed as not having any dominant word order in Dryer (2011r), but are listed as either VO or OV in Dryer (20110). This may be due to the fact that in these languages the position of the subject varies, but not that of the object. Macushi, for example, has both SOV and OVS word order with roughly the same frequency, which places it in the group of languages with no dominant order in Dryer (2011r). Nevertheless, the object fairly consistently precedes the verb, making it an OV language in Dryer's database on the position of the object with respect to the verb (20110).

Of the 189 languages listed in Dryer (2011r) as not having any dominant order, 39 are listed as having either SV or VS and OV or VO orders respectively in Dryer (2011q and 20110).[166] In these languages the subject tends to be in the same position relative to the verb, and the object tends to be in the same position relative to the verb, but the order of the subject and object relative to each other may vary. Of these, just under 60%

166. Cf. also Dryer (2011f).

are verb initial languages, which may indicate that variability of the position of S and O is somewhat more common in verb initial languages than in verb final or medial ones.

10.2.2 Order of phrase constituents

The constituents on the phrase level may also vary with respect to their order. This is especially true of NPs, where the order of the adjective and noun, genitive and noun, demonstrative and noun, numeral and noun, and so on may vary. Notice that for NP constituents we are, again, dealing with lexical constituents (i.e. those containing nouns) and not pronominal ones for the reasons discussed above.

In a genitive construction, or possessive construction, we have an NP consisting of a head noun (the **possessee** noun or the **possessum**) and the genitive noun phrase (the **possessor** phrase). The genitive phrase (possessor phrase) describers who or what the head noun belongs to. Examples of genitive constructions are *the man's hat*, *John's bike* and *the roof of the house*, where *the man's*, *John's* and *of the house* are genitive noun phrases (possessor phrases) and *hat*, *bike* and *the roof* are head nouns (possessee nouns). There are two logical possibilities for the order of constituents of genitive constructions, either that the genitive noun phrase precedes the head noun (abbreviated GenN) or that the genitive noun phrase follows the head noun (abbreviated NGen). The majority of languages in Dryer's (2011l) database, 685 of 1248 (or 54.9%), have a GenN order, while 467 (37.4%) languages have NGen and 96 (7.7%) do not have any dominant order. English belongs to the latter category, due to the fact that the genitive phrase either has the *'s* or *of NP* construction (as in *John's bike* and *the roof of the house*). Examples of the two logical orders are:

Evenki (Altaic (Tungusic): Russia)

(213) kungaka-r evike-r-tyn
child-PL toy-PL-3PL.POSS
Gen N
'(the) children's toys' (Nedjalkov 1997: 82)

Gela (Austronesian (Oceanic): Solomon Islands)

(214) mimi-na bolo
bladder-CONST pig
N Gen
'(the) pig's bladder' (Crowley 2002: 530)

The distribution of GenN and NGen is very similar to that of OV and VO respectively, except that GenN is more widespread than OV due to the fact that many SVO languages are GenN. (For a discussion on where the genitive marking appears in the genitive construction, see 7.2.1).

The pattern differs slightly for the APiCS languages, in that the two logical combinations are more evenly distributed than among the WALS languages (Michaelis et al. 2013: feature 2), where the proportions are hardly at all affected by the subtraction of the two contact languages in the sample, Ndyuka (with no dominant order) and Sango (with NGen order). Unlike in the WALS database, NGen as the dominant word order forms the biggest group in APiCS, with 35 of 76 languages (or 46.1%), while GenN is found in 33 languages (43.4%). The group of languages with no dominant word order is proportionally larger in APiCS as in WALS, with eight (10.5%) languages.

When it comes to NPs containing an adjective modifying the noun, there are again two logical word orders: either the adjective may precede the noun (AdjN) or it may follow the noun (NAdj). English is an example of an AdjN language: in *a red flower* the adjective (*red*) precedes the noun (*flower*). In Dryer's (2011h) sample of 1366 languages the absolute majority, 878 languages (or 64.3%), have NAdj as their dominating word order, while 373 (27.3%) have AdjN and 110 (8.1%) have no dominant word order. For five languages (0.4%), Choctaw, Mesa Grande Diegueño (Hokan (Yuman): USA), Kutenai, Seri and Jamul Tiipay, the closest equivalent to a modifying adjective is a type of relative clause (see further 11.2.3). Examples of the two logical orders are:

Gününa Küne (Chon (Puelche): Argentina)

(215) atek a bahai
mountain LK big
N Adj
'big mountain' (Adelaar & Muysken 2004: 562)

Assamese (Indo-European (Indic): India)

(216) dami kitap
costly book
Adj N
'costly books' (Goswami & Tamuli 2007: 475)

Languages with NAdj word order are spread over the world in Dryer's sample, but are almost entirely absent from the Eurasian mainland except for the far west of Europe and Southeast Asia, where AdjN word order dominates. Languages with AdjN word order also form a belt just under the Sahara and along the Northern American Pacific coast.

The pattern for the APiCS languages differs considerably in that the absolute majority of the languages (49 of 76 or 64.5%) have AdjN as their dominant word order while 20 (26.3%) have NAdj and seven languages (9.2%) have no dominant word order for modifying adjectives and the noun (Michaelis et al. 2013: feature 3).

A demonstrative can either take the form of a free word or an affix (see 6.2.2.1 for more on demonstratives), which can either precede or follow the noun. English has the demonstratives *this*/*that* and *these*/*those* (as in *this*/*that book* or *these*/*those books*). The logical combinations are thus that either the demonstrative word precedes the noun (DemN) or follows the noun (NDem). Or the demonstrative affix is either a prefix on the noun or a suffix on the noun. Lastly, the demonstrative word or affix may simultaneously precede and follow the noun. In Dryer (2011k) 560 languages (45.8%) have NDem as their dominant order, just slightly more than the 542 languages of 1223 (44.3%) with DemN as their dominant order. Examples (218) and (217) show the two orders:

Konzime (Niger-Congo (Bantoid): Cameroon)

(217) mʉd ɲâ(k)
person this.CL1
N Dem
'this person' (Beavon 1986: 174)

Minnesota Ojibwe (Algic (Algonquian): USA)

(218) a'aw wemittokōs̈i
that Frenchman
Dem N
'that Frenchman' (Schwartz & Dunnigan 1986: 300)

Demonstrative affixes are uncommon in the sample, with only nine (0.7%) languages with a demonstrative prefix and 28 (2.3%) with a demonstrative suffix. Barbareño Chumash is an example of a language with a demonstrative prefix (*heʔ-* 'this') and Agar Dinka is an example of a language with a demonstrative suffix (*-è̱* 'this').

Barbareño Chumash (Chumash (Chumash): USA)

(219) heʔ-l-ʔašḵaʔ
this-ART-coyote
Dem-N
'this coyote' (Wash 2001: 58)

Agar Dinka (Nilo-Saharan (Nilotic): Sudan)

(220) mo̱ŋ-è̱
man.CS1-DEM1
N-Dem
'this man' (Andersen 2002: 14)

To have a simultaneous demonstrative word or affix both preceding and following the noun is also rare, found in 17 languages (1.4%). An example of such a language is Malagasy.

Malagasy (Austronesian (Barito): Madagascar)

(221) io trano io
this house this
Dem N Dem
'this house' (Keenan & Polinsky 1998: 567)

Note that both elements in (221) need to flank the noun for the phrase to be grammatical; something like **io trano* or **trano io* is not acceptable. For 67 languages (5.5%) no order or strategy is dominant. Languages where the demonstrative precedes the noun (whether as a free word or a prefix) are very common on the Eurasian mainland in Dryer's sample, except for the far west of Europe and Southeast Asia. They are also very common in both North and South America. Languages where the demonstrative follows the noun are particularly common in Africa, Southeast Asia and Papua New Guinea.

The figures for the languages in APiCS are at the time of writing still temporary for this feature, but they seem to differ somewhat in their pattern from those in WALS (where the pattern is not noticeably affected by the subtraction of the two contact languages in the sample, Ndyuka and Sango, both NDem). More than half seem to have DemN as their dominant order, while only just under at third have NDem (Michaelis et al. 2013: feature 5). Only very few seem to lack any dominant word order for demonstratives and even fewer seem to have any demonstrative affix.

Languages also differ with respect to the relative position of the numeral and the noun in such phrases where the number of things are referred to (for which cardinal numbers are used, as opposed to when the rank in the order of items is referred to, for which ordinal numbers are used). Of the 1154 languages in Dryer's (2011n) sample, the majority, 608 (or 52.7%), have the numeral following the noun (NNum) as their dominant order, while in 479 languages (41.5%) the numeral preceding the noun is the dominant order and 65 languages (5.6%) have no dominant order. The languages in Dryer's sample with NNum as their dominant order predominate in sub-Saharan Africa and a limited area of Southeast Asia stretching roughly from Nepal to Cambodia and Papua New Guinea. Europe, northern Africa, the Middle East and Asia are almost entirely dominated by languages with NumN order, except for the aforementioned area in Southeast Asia. NumN languages also predominate along the Pacific coast of North America and North-western South America. English is an example of the latter, as in *five books*. Other examples are:

Buma (Austronesian (Oceanic): Solomon Islands)

(222) mʷoe tili
house five
N Num
'five houses' (Tryon 2002: 576)

Hungarian (Uralic (Ugric): Hungary)

(223) négy toll
four pen
Num N
'four pens' (Kenesei et al. 1998: 96)

APPROXIMATION THROUGH WORD ORDER

East Slavic languages have the typologically very rare strategy of expressing approximation ('about X, roughly X') by inverting the word order of numeral and noun:

Russian (Indo-European (Slavic): Russia)

a. na dva mal'čika
PRQ two boy.SG.GEN
'two boys'

b. mal'čika na dva
boy.SG.GEN PRQ two
'about two boys' (Billings 1995: 31, 33)

In two languages in Dryer's sample (0.2%), Wari' and Arára Karó, the only normal way of expressing a numeral is a construction where the numeral does not modify the noun but the verb.

Arára Karó (Tupian (Ramarama): Brazil)

(224) maʔwɨt ip ʔɨy matet cagárokōm=nem
man fish catch.IND yesterday two=ADVZ
'The man caught two fish yesterday.' (Dryer 2011n citing Gabas, Jr 1999: 172)

In (224) the numeral takes the form and position of an adverb in a construction which is similar to the English sentence *The men have both gone*, where *both* takes the position of an adverbial.

The figures for the languages in APiCS are at the time of writing still temporary for this feature, but they seem to indicate that pattern among in the APiCS sample differs a great deal from that in WALS (where the subtraction of Ndyuka (NumN) and Sango (NNum) does not affect the pattern noticeably). In APiCS it seems as if almost nine tenths have NumN as their dominant order, while only a few languages have NNum as their dominant order (Michaelis et al. 2013: feature 6).

There seems to be a certain correlation between the various word orders discussed above, in that some modifiers seem to parallel each other with respect to word order. For instance, the majority of the AdjN languages in the printed version of WALS (Haspelmath et al. 2005) are also DemN, and the majority of DemN languages are also NumN. Conversely, slightly more than half of the NAdj languages are also NDem and the majority of NDem languages are also NNum. This is, however, by no means any set rule and is only a rather weak tendency. Furthermore, the tendency only holds on a binary basis, that is, if we combine two features with each other. Thus, if we combine all features of GenN + AdjN + DemN + NumN (that is, all those instances discussed above where the modifier[167] precedes the noun) we get 144 languages, and if we combine the features NGen + NAdj + NDem + NNum (that is, all those instances where the modifier follows the noun) we get 116 languages, the sum of which (260) is only about a quarter of the languages in Dryer's (2005a, 2005b, 2005c and 2005d) maps.

10.2.3 Constituent order correlations

The position of the verb tends to correlate with a number of other word orders on different levels. This is especially true when it comes to the position of the verb with respect to the object. OV languages typically behave in one way while VO languages typically behave in a different way. The following will list some, but by no means all, typical characteristics of OV and VO languages respectively. I stress, however, that these are tendencies and not absolute universals. For a very detailed discussion on word order correlations, see Dryer (2007b). It is beyond the scope of this chapter to discuss the various attempts at explanations for why languages exhibit the word order patterns and correlations that they do. For a very accessible overview of the various theories that have been put forth, see Song (2009). See also Dryer (2007b:130f) for further references.

10.2.3.1 *OV languages*

Most OV languages are verb final languages, i.e. languages where the verb follows both the subject and the object, except for the rare OVS languages, which are essentially verb medial.

An example of a fairly typical OV language is Khwarshi, which displays a number of characteristics with respect to constituent order that are commonly found in OV languages. First of all, obviously the object precedes the verb.

---

167. For the sake of simplicity I am here treating a genitive construction as a modifier plus noun, along the lines that the genitive phrase provides specifying or descriptive information about the noun, much in the same way as a demonstrative does.

Khwarshi (Nakh-Daghestanian (Avar-Andic-Tsezic): Russia)

(225) obu-ť'-i łido b-łť'-x-i
father-OBL-ERG firewood(III) III-divide-CAUS-PAST.W
O V
'The father chopped the firewood.' (Khalilova 2009: 69)

In OV languages, the adverb typically precedes the verb:

(226) xexiłin Muħamad ø-oť'uq'-še golło
fast Magomed(I) I-come-PRES be.PRES.COND
Adv V
'If only Magomed would come quickly.' (Khalilova 2009: 252)

In (226) the adverb *xexiłin* 'fast, quickly' is placed prior to the verb *øot'uqše* 'come. PRESent tense' and its auxiliary *golło* 'would'.

Khwarshi is also typical in that the genitive phrase (possessor phrase) precedes the head noun (the possessee or possessum):

(227) obu-ť'-łs tubi
father-OBL-GEN1 gun
Gen N
'father's gun' (Khalilova 2009: 70)

When comparisons are made, we need a standard of comparison (St), a marker of comparison (M) and an adjective (Adj), as in *Mary is taller than Jane*, where *Jane* is the standard of comparison (she is what we are comparing with), *than* is the marker of comparison (i.e. the marker indicating that X is being compared with Y) and *taller* is the adjective of comparison. In OV languages the standard of comparison is typically followed by the marker of comparison, which is in turn followed by the adjective:

(228) Qurban Nazir-λ'o-zi-n lebala-w goli
Kurban(I) Nazir-SUP-ABL-and brave-I be.PRES
St M Adj
'Kurban is braver than Nazir.' (Khalilova 2009: 76)

In (228) we are comparing Kurban with Nazir, so Nazir is the standard of comparison. The marker, indicating that comparison is being made, is here the suffix *-λ'o*, which attaches to the standard of comparison (Nazir). The adjective of comparison, *lebalaw* 'brave', follows the construction standard-marker (*Nazir-λ'o-*).

In OV languages adpositions tend to follow the noun, i.e. are postpositions, as in (229) where *λ'olo* 'above' is placed after the noun *yoboλ'o* 'pile.Case marker':

(229) ɣobo-λ'o λ'olo ɣur-a-ba l-eč-i
pile-SUP above stone-OBL-PL.ABS NHPL-be-PAST.W
N Po
'There were stones on the top of the pile.' (Khalilova 2009: 135)

That an OV language like Khwarshi has postpositions is, in fact, a bidirectional implicational universal (cf. 2.3.2). That is, the correlation between OV languages and postpositions is so strong that we may claim that a postpositional language is an OV language, and, conversely, if the language is OV then it will have postpositions. Other examples of bidirectional correlations with OV languages are that the adpositional phrase precedes the verb and that the auxiliary follows the verb.

(230) [miɬɬo b-oλoλ'o] henše gul-o
[2PL.GEN2 III-in.middle] book(IIII) put-IMP
PP V
'Put the book between you!' (Khalilova 2009: 42)

In (230) the adpositional phrase *miɬɬo boλoλ'o* 'between you.PLural' precedes the verb *gulo* 'put'.

(231) íljo b-odo-še goli
1PL.ABS HPL-work-PRES be.PRES
V Aux
'We are working.' (Khalilova 2009: 188)

In (231) the auxiliary *goli* 'are' follows the lexical verb *bodoše* 'working'.

Khwarshi is also typical for an OV language in that it does not have articles. However, while the majority of OV languages lack articles, it is a bidirectional correlation that if an OV language has articles, then the article will follow the noun, and conversely, if in a language the article follows the noun, then it is highly likely to be an OV language. Mugil is an example of an OV language that has articles:

Mugil (Trans-New Guinea (Madang): Papua New Guinea)

(232) in karuw araʔ yag-o1[168]-iy
they animal a give.me-PAST-N1P
N (O) Art V
'They gave me a pig.' (Hepner 2006: 28)

168. The notation o1 indicates a place holder for a zero morpheme indicating either tense or mood (Hepner 2006: 11).

In (232) the object, the noun *karuw* 'animal', precedes the verb, *yagiy* 'gave', so it is an OV language. Furthermore, the article *araʔ* 'a' follows the noun it classifies (*karuw*). Mugil thus conforms to the typical pattern of OV languages with articles.

10.2.3.2 *VO languages*

VO languages comprise both the verb initial languages with VSO and VOS word order, and the large group of verb medial languages with SVO order. They are in many ways the mirror image of OV languages. SVO languages behave almost exactly like verb initial languages, except when it comes to the order of the genitive phrase and the head noun, as we will see below. To illustrate the similarities between verb initial languages and SVO languages, I will give examples of both a verb initial language and an SVO language; unless otherwise stated, the following examples are from San Bartholomé Zoogocho Zapotec and English.

San Bartholomé Zoogocho Zapotec (Oto-Manguean (Zapotecan): Mexico)

(233) dx-aogo be'ko' yet
CONT-eat dog tortilla
V O
'The dog is eating tortillas.' (Sonnenschein 2004: 126)

(234) The dog is eating tortillas.
V O

As Examples (233) and (234) show, irrespective of whether the verb is sentence initial or not, the relative order between the verb and the object is the same for both languages.

In VO languages, the adverb tends to follow the verb:

(235) ble=na' dx-lonhgh=e' sholazhe
CLAN=DEMDIST CONT-run=3FM slowly
V Adv
'That person runs slowly.' (Sonnenschein 2004: 161)

(236) That person runs slowly.
V Adv

It should be noted, however, that the position of the adverb is relatively free in San Bartholomé Zoogocho Zapotec (Sonnenschein 2004: 251; adverbs may also vary some in English). An example of a VO language where the adverb has a more rigid order is Fijian, where the manner adverb follows the verb (Dixon 1988: 63).

Fijian (Austronesian (Oceanic): Fiji)

(237) bau 'ada va'a.totolo noo
MODIF run quickly ASP
V Adv
'Try and run more quickly.' (Dixon 1988: 80)

In VO languages the adpositions tend to precede the noun (or noun phrase), i.e. are prepositions:

(238) lenh yaa wag
with iron wood
Pr NP
'with an axe' (Sonnenschein 2004: 332)

(239) with an axe
Pr NP

In genitive constructions verb initial languages tend to have the order NGen. However, with SVO languages, both NGen and GenN orders are common. English is an example of a language with no dominant order for genitive constructions (cf. Example (241)), but for most SVO languages a dominant order of either NGen or GenN can be identified.

(240) yichgh bedw
head PN
N Gen
'Pedro's head.' (Sonnenschein 2004: 127)

(241) Pedro's head ~ the roof of the house
Gen N N Gen

In comparative constructions in VO languages, the marker of comparison tends to follow the adjective, and the standard of comparison tends to follow the marker of comparison:

(242) n-ak=dx bdxee be lis kleka' be'ko'
STAT-be=more ant CLAN small than dog
Adj M St
'Ants are smaller than dogs.' (Sonnenschein 2004: 232)

(243) Ants are smaller than dogs.
Adj M St

In (242) the adjective *lis* 'small' precedes the marker of comparison *kleka'*, which in turn precedes the standard of comparison (the thing we are comparing with), *be'ko'* 'dog'. As (243) shows, English exhibits the same pattern. With the bidirectional correlations, again VO languages are essentially the mirror image of OV languages. Thus the adpositional phrase follows the verb, the auxiliary precedes the verb, and the article precedes the noun.

(244) sh-chog=a' yag-en [lenh yaa wag]
CONT-cut=1SG tree-DET [with iron wood]
V PP
'I cut the tree with an axe.' (Sonnenschein 2004: 332)

(245) I cut the tree [with an axe].
V PP

In (245) the prepositional phrase *lehn yaa wag* 'with an axe' is placed to the right of the verb; English exhibits the same pattern.

In VO languages the auxiliaries, if and when used, tend to be placed before the lexical verb:

(246) dx-eyala si-i=e' shon gayoa
CONT-should POT-grab=3F three hundred
Aux V
'She should grab three hundred.' (Sonnenschein 2004: 216)

(247) She should grab three hundred.
Aux V

Finally, articles tend to precede the noun in VO languages:

(248) b-enh bidao neda' to libr
COMPL-give child 1SG one book
Art N
'The child gave me a book.' (Sonnenschein 2004: 157)

(249) The child gave me a book.
Art N

10.2.4 Flexible word order

A number of languages have so-called 'flexible' (or 'free') word orders, where the order of the constituents is not determined by syntactic roles, such as subject and object, but by some other principle. In some languages the order is determined by semantic factors such as animacy, in others by various pragmatic factors such as topicality or focus, or information structure, in yet others the word order is determined by grammatical factors such as aspect. In Navajo, for example, the order of the subject and

object is dependent on the relevant NP's position on the animacy hierarchy according to the following principles:

human > animals (large > medium > small) > insects > natural forces > plants & inanimate objects > abstract notions

The ranking thus not only takes animacy into account, but, with animals, also size. In other words, a bigger animal ranks higher than a smaller one, a small animal ranks higher than an insect, and so on. The NP ranking higher on the hierarchy appears further to the left in the sentence.

Navajo (Na-Dene (Athapaskan): USA)

(250) a. dóola dibé yiyiisxį́
bull sheep killed
S O V
'The bull killed the sheep.'

b. *dibé dóola biyiisxį́
sheep bull killed
(Uyechi 1996: 124)

c. *toís'ná na'astǫǫsí yishish
bee mouse stung
'The bee stung the mouse.'

d. na'astǫǫsí toís'ná bishish
mouse bee stung
O S V
(Uyechi 1996: 124)

In (250) the order of the subject and object relative to each other are not dependent on grammatical relations, but on position on the animacy hierarchy. Thus, (250a) is grammatical, since the bull is bigger than the sheep, but (250c) is not, since the mouse is higher up on the hierarchy than the bee. The mouse therefore has to precede the bee in word order, as in (250d), irrespective of what syntactic role it has.[169]

Cayuga is an example of a language where the object either precedes of follows the verb, depending on pragmatic factors.

Cayuga (Iroquoian (Northern Iroquoian): USA)

(251) a. katsihwá' kihsa:s
hammer 1SG.seek
O V
'I'm looking for a hammer.' (said in shop, with no particular hammer in mind)

b. to: ti' nika:nô:' nê:kyê katsíhwa'
how then so.it.costs this hammer
V O
'How much does this hammer cost?' (indicating specific hammer)
(Mithun 1992: 28)

169. The person marking on the verb alternates between *yi-* and *bi-* depending on whether or not the word order has been 'inverted' (i.e. with the object placed before the subject). See further Uyechi (1996).

In (251a) the object is a non-specific newly introduced piece of information or item and thus comes early in the clause. In (251b) the object is a specifically identified item and thus appears later in the clause.

In Yi the word order varies according to the grammatical aspect of the proposition. In clauses denoting ongoing events, the order is SOV, while in clauses denoting resultative events, the order is OSV.

Yi (Sino-Tibetan (Burmese-Lolo): China)

(252) a. m̩^{33}ka^{55} ʂa^{33}ma^{55} kᵛũ̯^{33} ȵɛ̝dʑɔ^{33}
PN PN frighten CONT
S O V
'Muga is frightening Shama.' (ongoing)

b. lo^{55}ʈɕi^{33} ʈɕi^{33} ŋa^{33} ɖʑɔ^{33} ko^{44}ʂa^{33} o^{44}
finger CL 1SG fell SEND DP
O S V
'I have cut off my finger.' (resultative) (Gerner 2004: 114, 117)

In (252a) the event is ongoing, and the subject precedes the object. In (252b), however, we have an event in the resultative, and the subject follows the object.

In these kinds of languages it is only accurate to label it 'flexible' word order if the criterion is the position of constituents according to grammatical relations, since they are only flexible in the sense that constituents with various syntactic roles may appear in different positions. However, the word order in the above-mentioned languages may be said to be rigid with respect to other criteria, such as animacy, pragmatic roles or whatever else it is that determines the position of the constituents.

10.3 Word order in sign languages

We have seen that it is less than straightforward to determine the basic word order for spoken languages. In a sense, it is even less straightforward for signed languages. This is partly because it is less common in signed languages than in spoken languages to have clauses containing two overt nominal arguments; in fact it seems to be very common indeed for all known sign languages to allow clauses that only contain the verb, where the arguments are covert and only implied through context. This is especially true for clauses with pronominal arguments, but also nominal arguments may be covert, as in Example (253). Some sign languages, such as IPSL, avoid clauses with more than one nominal argument altogether (Zeshan 2003c).

AdaSL (Sign Language: Ghana)

(253) MOVE FLY-TO-SIDE BALL TAKE:CL-round LIFT&THROW:CL-round ENTER
'(The bird) flew to the side, took (a) ball and threw it inside.' (Nyst 2007: 165)

In (253) the subject (*the bird*) is not overtly expressed, but has to be inferred from the context.

With pronouns it is common for a clause to lack both arguments, since the verb may agree with both the subject and the object through direction. The pronouns are indicated with subscript number, where 1 means 'fist person' and 2 'second person'.

LIU (Sign Language: Jordan)

(254) $_1$TELL$_2$
'I tell you.' (Hendriks 2008: 59)

In (254) the sign for TELL involves the motion from the subject (whoever does the telling) towards the object (whoever is being told). Reversing the sign to $_2$TELL$_1$ would therefore reverse the meaning to 'You tell me'. In both cases the arguments are only marked through agreement on the verb (the direction of the motion) and are not overtly signed.

Nevertheless, while the word order is typically less rigid in signed languages, a basic word order has been determined for a number of sign languages. ST, for example, is SVO, while DGS is SOV.

ST (Sign Language: Sweden)

(255) DRIVER WASH CAR
S V O
'The driver washed the car.' (Ahlgren & Bergman 2006: 43)

DGS (Sign Language: Germany)

(256) MAN BOOK BUY
S O V
'The man buys a book.' (source: personal knowledge)

While there are no large-scale surveys of the word order in sign languages yet, Map 10.1 gives an overview of 30 sign languages. Notice that the word order of the sign language is independent from the word order of the spoken language in the area. Thus BSL has no dominant order, while spoken English has SVO as its dominant order, and while LSA in Argentina has SOV word order, spoken Spanish has SVO as its dominant word order.

Map 10.1 Word order in 30 sign languages. Black dots: SVO (12 languages); black squares: SOV (5 languages); black triangles: OSV (2 languages); white dots: flexible word order (11 languages). For a full legend, see http://dx.doi.org/10.1075/z.176.additional.

The most common word order reported is SVO, while SOV – the most common word order for spoken languages – has only been reported for five languages: ABSL, LSA, ÖGS, DGS and NS. OSV has been reported as the basic word order in CSL (Massone & Curiel 2004), and possibly also in UKSL (Kakumasu 1968). A much higher proportion of sign languages than spoken languages have a flexible word order based on pragmatic factors such as topic and comment. Most typically the topic of the proposition is signed first.

It will be a matter for future research to investigate which word order correlations can be established for sign languages and to what extent this mirrors or differs from the correlations found for spoken languages.

10.4 Summary

Languages tend to organize their utterances structurally according to some principle. Syntax investigates how meaningful units, or constituents, group together to form larger units, which in turn group together to form grammatical sentences. Constituents may be arranged in various ways, and some languages are more rigid in which kinds of arrangements they allow than others.

Word order typology investigates the various ways constituents may be ordered and how the various orders pattern across the world. When investigating the dominant constituent order on any given level, whether phrasal or clausal, the focus is on unmarked declarative sentences where the argument or arguments of the verb are nominal and not pronominal.

On the clause level the patterns may either differ greatly or be rather balanced depending on how many and which constituents are being mapped. Thus the vast majority of languages are SV and not VS, while the orders VO and OV are equally common. For transitive clauses, the majority of languages are either SOV or SVO, but all six logical possibilities are attested in the world.

Phrase constituents may also vary with respect to word order. There tends to be a correlation between clausal and phrasal constituent orders, especially between the order of the object and the verb and various phrasal word orders.

Some languages have so-called flexible word order in that they organize their constituents according to other criteria than grammatical relations, such as pragmatic roles or animacy.

Sign languages also vary with respect to their constituent orders. SVO seems rather common but others are attested. However, most sign languages are more flexible in their word order than spoken languages, and with many the word order is determined by pragmatic factors.

10.5 Keywords

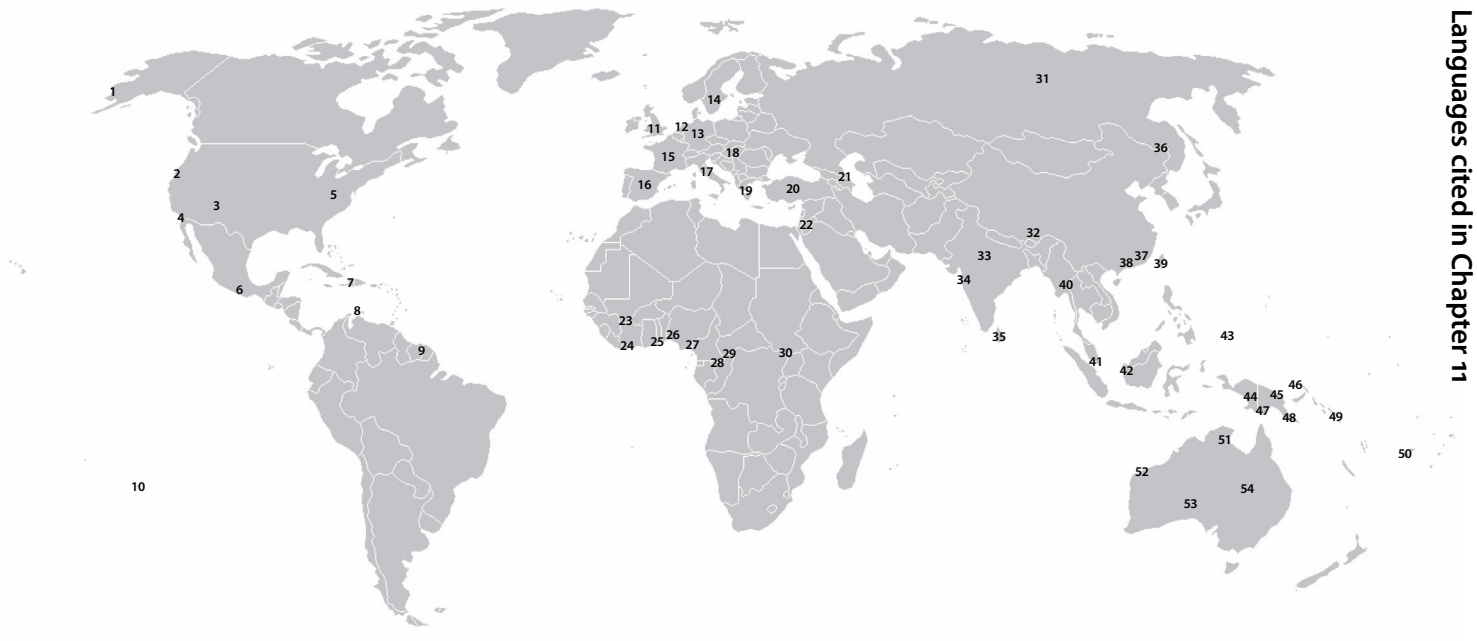
basic word order
clause
constituent
constituent order
constituent order correlations
flexible word order
OV languages
phrase
sentence
VO languages

10.6 Exercises

1. What kinds of sentences are thought to reflect the basic word order of a language? Explain why.
2. How do pidgin and creole languages pattern with respect to phrasal word order compared to non-creole languages? How does the pattern correlate (if at all) with major lexifiers?
3. Describe the typical constituent order patterns found in VO and OV languages respectively.
4. How do sign languages compare with spoken languages with respect to establishing basic word order?
5. Is the following statement true or false? Motivate your answer.

   In languages with flexible word order constituents are arranged haphazardly.

Languages cited in Chapter 11



1 Yup'ik, Central
2 Pomo, Eastern
3 Acoma
4 Tiipay, Jamul
5 American Sign Language
6 Mixtec, Alacatlatzala
7 Haitian Creole
8 Papiamentu
9 Ndyuka
10 Rapanui
11 English
12 Dutch Sign Language
13 German
14 Swedish
Swedish Sign Language
15 French
16 Spanish
17 Italian Sign Language
Latin
18 Hungarian
19 Greek, Homeric
20 Turkish
21 Dargwa
Khwarshi
Lezgian
22 Jordanian Sign Language
23 Supyire
24 Godié
25 Adamorobe Sign Language
Logba
26 Yoruba
27 Babungo
Lokaa
Ngemba
28 Baka
29 Sango
30 Logoti
31 Evenki
32 Tibetan, Classical
33 Indo-Pakistani Sign Language
34 Korlai
35 Sri Lankan Malay
Sri Lanka Portuguese
36 Chinese Pidgin Russian
37 Hakka
38 Cantonese
39 Amis
40 Kayah Li, Eastern
41 Singapore Bazaar Malay
42 Mualang
43 Sonsorol-Tobi
44 Kombai
45 Amele
Mauwake
Yimas-Arafundi Pidgin
46 Loniu
47 Kamula
48 Barai
49 Toqabaqita
50 Pidgin Hindustani
51 Kriol
52 Martuthunira
53 Pitjantjatjara
54 Australian Sign Language

Chapter 11

Complex clauses

We have seen that languages have ways of combining elements into larger units, where the elements take their various roles within the larger unit. Languages may also combine elements to jointly take a certain role in a larger unit. Combined units may be of the same or of different ranks. Section 11.1 gives an overview of the main features of coordinated clauses, where units of the same rank are linked together to form a complex clause. In Section 11.2 the three main types of subordinated clauses are discussed: those that serve as arguments to their matrix (11.2.1), those that serve as adjuncts to their matrix (11.2.2) and those that serve as NP modifiers (11.2.3). Section 11.3 briefly summarizes the notion of cosubordination, where clauses neither are embedded in each other nor can function independently. This specifically relates to serial verb constructions (11.3.1) and clause chaining (11.3.2). Finally, Section 11.4 sketches some of the characteristics of complex clauses in sign languages.

11.1 Coordination

Coordination is when linguistic units of the same syntactic status are linked together to form a larger linguistic unit which carries the same function as its parts. For instance, a clause may have a subject, a verb and an object, as in *The man boiled the water*. However, we may combine two verbs and still get a clause with a subject, verb and object, as in *The man boiled and poured the water*, where both *boiled* and *poured* combine on the same level to jointly carry out the same function of the clause predicate. This may also be done with phrases, as in *The man and the woman boiled the water*, where *the man* and *the woman* combine on the same level, or it may be done with full clauses, as in *The man boiled the water and the woman filled the teapot*, where *the man boiled the water* and *the woman filled the teapot* combine on the same level. What is crucial here is that none of the combined units are dependent on each other. It is not essential for the verb *boil* to be combined with the verb *pour* in order to make a well-formed clause, nor is it essential for the NP *the man* to combine with the NP *the woman* in order to make the clause complete, nor for the two clauses *The man boiled the water* plus *The woman filled the teapot* to be combined in order to make a grammatical sentence. In other words, independent units are coordinated with each other.

All known languages have some way of combining elements of the same levels into larger units where they jointly have a certain function, but the strategies may differ. Units may be coordinated by simply placing them next to each other (juxtaposition), as in *I came in, sat down, crossed my legs*. Or an overt linker may be used, a **coordinator** (also called **coordinating conjunction**).[170] The most frequently used coordinator is the **conjunctive coordination** (or **conjunction**) *and* (as in *I came in and sat down*). Other coordinators are the **disjunctive coordination** (or **disjunction**) *or* (as in *I wanted an apple or a pear*), the **adversative coordination** *but* (*I was tired but happy*), and the **causal coordination** *for* (as in *I fell asleep, for I was exhausted*). For an accessible typological overview of coordination, see Haspelmath (2007), which this section relies heavily on.

11.1.1 Asyndetic coordination

Asyndetic coordination, or **juxtaposition** (also called the '**zero strategy**', cf. Payne 1985b: 25ff), is the coordination of units without any overt linker. The term itself derives from Greek *a-sýndetos* 'not-bound.together'. This is a very common strategy in the world. Many languages permit juxtaposition even though it has coordinators, as is the case in English: compare the sentences *Quietly and furtively the detective searched the desk of the suspect* and *Quietly, furtively the detective searched the desk of the suspect* where the coordinator (underlined) in the first sentence is simply omitted in the second. In English juxtaposition is stylistically marked and is generally used to achieve specific effects in written language. Juxtaposition is also possible for clauses in English, although it also tends to be stylistically marked. Compare the sentence *John saw Mary and Mary smiled at him* with *John saw Mary. Mary smiled at him*. Notice that with juxtaposition intonation is the only way to signal that we have a coordinated construction (illustrated in the English examples with punctuation, such as commas or full stops), that is, intonation is the only indicator that the units we are linking belong together in a larger unit.

In many languages juxtaposition is the default method of coordination, although they may have optional overt linkers. Rapanui is an example of a language where coordination is normally expressed through juxtaposition, both for words, phrases and clauses:

---

170. Due to the similarity between the terms *coordinating conjunction* and *conjunctive coordination*, I am here following Haspelmath (2007) and using the term **coordinator** to indicate the overt linker of units in coordinating constructions.

Rapanui (Austronesian (Oceanic): Easter Island)

(257) a. me'e nehe nehe iti iti era ko ia
thing pretty RED little RED PPD FOC 3SG
'She was pretty (and) small.'

b. he e'a kiruŋa, he u'i u'i a te kona ta'ato'a
ACTN get up ACTN look RED POSS SPE place every
'She stood up (and) looked everywhere.'

c. he oho he tu'u he u'i, i oho era, i u'i era he ki
ACTN go ACTN arrive ACT see PAST go PPD PAST see PPD ACTN say
mai ki a au
TOW DAT P.SG 1SG
'I went off (and) came to see him, (and) when I had got there, (and) he had seen me, he said to me…' (Du Feu 1996: 85, 88)

In (257a) the adjectives *nehe nehe* 'pretty (reduplicated)' and *iti iti* 'small (reduplicated)' are coordinated; they are simply juxtaposed. In (257b) the verb phrases (*he*) *a'a kiruŋa* 'get up' and (*he*) *u'i* 'look' are also simply juxtaposed, as are the coordinated clauses in (257c), indicated in the example by commas. This is also the default strategy for disjunctive coordinations ('or'-coordination):

Rapanui (Austronesian (Oceanic): Easter Island)

(258) he oro ki te po'e mo kai, he inaki koe ki te kiko
ACTN grate DAT SPE po'e BEN eat ACTN accompany 2SG DAT SPE meat
'You grate it to make po'e (or) you use it to accompany meat.' (Du Feu 1996: 87)

In (258) the two clauses are disjunctively coordinated through juxtaposition only, indicated by a comma.[171]

171. It should be noted that conjunctive and disjunctive coordination may in Rapanui optionally be marked with a linker borrowed from Spanish. It is very common for languages with no coordinators that have been colonized to adopt the coordinators of the colonizing language. Very often this is due to the bilingual setting of societies that were colonized, and it seems reasonable to assume that one pulling factor for borrowing coordinators is the emerging literacy: many languages with overt coordinators have a long literary tradition – coordination by juxtaposition is dependent on intonation, which does not show in written language. For languages with a young literary tradition the need for overt coordination markers is thus recent and a convenient way of filling that need is to simply import the necessary marker. The colonizing language is then typically the easiest available source. See further Mithun (1988). For more on linguistic borrowing, see 13.2.

11.1.2 Syndetic coordination

Syndetic coordination (from Greek *sýndetos* 'bound.together'), sometimes also called **linked coordination**, involves an overt linking device, or coordinator (often called conjunction). The coordinators may either precede the unit they are coordinating (**prepositive** coordinators) or may follow the units they are coordinating (**postpositive** coordinators), and they may be either free words or bound clitics. Furthermore, if two units are being coordinated, this may be done with only one coordinator, as in English (*A and B*) or with two coordinators (something like *A and B and*). Coordination done with only one coordinator may be termed **monosyndetic**, while **bisyndetic** coordination involves two overt linkers. Each strategy will be exemplified below. For the sake of simplicity, I will here limit the discussion to binary coordinations only, that is, when only two elements are coordinated. Following Haspelmath (2007) I will label the overt linker *co* (for coordinator) and the two units that get coordinated *A* and *B* respectively.

Due to space limitations, the focus will be on conjunctions. In English the conjunction appears in the same position as other coordinators, such as disjunctions (compare *pizza and pasta* with *pizza or pasta*). This need not be the case. In Khwarshi, for example, while conjunctions and disjunctions are both bisyndetic, the former is postpositive while the latter is prepositive:

Khwarshi (Nakh-Daghestanian (Avar-Andic-Tsezic): Russia)

(259) a. bataxu-**n** k'oro-**n**
bread.V-and cheese.IV-and
'bread and cheese'

b. **ya** kad **ya** uže
or girl or boy
'girl or boy' (Khalilova 2009: 311f)

As mentioned in 6.2.2.6, languages differ as to whether they have the same coordinator for NPs as they do in VPs and clauses. Languages also differ as to whether or not the conjunction is the same word as the comitative ('with'). In English the NP conjunction differs from the comitative: *The man and the woman went to the market* (NP conjunction) versus *The man went to the market with the woman* (comitative phrase). This is slightly more common than to use the same marker for the two constructions: of the 234 languages in Stassen's (2011) database, 131 (or 56%) differentiate between conjunctions and comitatives, while 103 (or 44%) do not. An example of a language where the two are not differentiated is Babungo:

Babungo (Niger-Congo (Bantoid): Cameroon)

(260) Làmbí nə̀ Ndùlá gə̀ táa yìwìŋ
PN and/with PN go.PFV to market
'Lambi and Ndula went to the market' / 'Lambi went to the market with Ndula.'
(Schaub 1985: 84)

The proportions in the WALS sample do not change with the subtraction of the two contact languages in the sample, Haitian Creole and Sango, none of which differentiate between 'and' and 'with'. The figures for the languages in APiCS are at the time of writing still temporary for this feature, but they seem to exhibit a similar pattern, in that differentiation is somewhat more common than not (Michaelis et al. 2013: feature 71). In about a fifth of the languages there seems to be an overlap between the two markers in that one marker may function both as a conjunction and as a comitative, while the other has only one of those functions. Papiamentu is an example of such a language:

Papiamentu (Creole (Spanish-lexified): Netherlands Antilles)

(261) a. ela kumpra pan **i** keshi
3SG buy.PFV bread and cheese
'He bought bread and cheese.'

b. ela kumpra pan **ku** keshi
3SG buy.PFV bread with/and cheese
'He bought a cheese sandwich.' Or: 'He bought bread and cheese.'

(Muller 1989: 472)

The coordinator *i* in (261a) can only function as a conjunction ('and'), while the marker *ku* 'with/and' in (261b) may function both as a comitative ('with') and as a conjunction ('and').

11.1.2.1 *Monosyndetic coordination*

In monosyndetic coordination we have only one overt linking device. There are four logical possibilities for monosyndetic coordination: prepositive with the coordinator preceding the first unit (*co(-)A B*); prepositive with the coordinator preceding the second unit (*A co(-)B*, also called medial prepositive); postpositive with the coordinator following the first unit (*A(-)co B*, also called medial postpositive); and postpositive with the coordinator following the second unit (*A B(-)co*).

The most common strategy for coordination is *A co(-)B*, i.e. where the coordinator precedes the second coordinated unit (cf. Stassen 2000, who terms this strategy **medial monosyndeton**). English is an example of such a language; one indicator for the fact that the coordinator groups with the second unit in a phrase like *John and Peter* is prosody. If we make a pause between the two units, that pause tends to come before the coordinator (compare *John… and Peter* with ??*John and… Peter*). Another indicator is where the coordinator ends up if the two units get separated: compare *John will come to the party, and Peter if he can make it* with **John and will come to the party, Peter if he can make it*.

Postpositive monosyndetic coordination is cross-linguistically rather rare (Stassen 2011). An example of the *A(-)co B* strategy, i.e. where the coordinator follows the first unit, can be found in Lezgian:

Lezgian (Nakh-Daghestanian (Lezgic): Russia)

(262) Ali.di-ni-ni Weli.di-n buba
PN-GEN-and PN-GEN father
'Ali's and Weli's father' (Haspelmath 1993: 327)

In (262) the coordinating suffix *-ni* 'and' attaches to the first of the two coordinated units.[172] An example of the *A B(-)co* strategy, i.e. where the coordinator follows the second unit, can be found in Pitjantjatjara:

Pitjantjatjara (Australian (Pama-Nyungan): Australia)

(263) Henry-ku mama ngunytju puru
PN-GEN mother father and
'Henry's father and mother' (Stassen 2000: 15 citing Glass & Hackett 1970: 66)

In (263) the coordinator *puru* 'and' follows the second of the two coordinated units.

The fourth logical possibility, *co(-)A B*, i.e. where the coordinator precedes the first of the two coordinated elements, is not currently known to exist (at least not for conjunctions).

11.1.2.2 *Bisyndetic coordination*

With bisyndetic coordination there are also four logical possible strategies: that both coordinators are prepositive (*co(-)A co(-)B*), that both coordinators are postpositive (*A(-)co B(-)co*) or that the coordinators are mixed with respect to position (either *co(-)A B(-)co* or *A(-)co co(-)B*). All four types are attested, but not all are the default strategy for coordination. Prepositive bisyndetic coordination (*co(-)A co(-)B*), for example, only occurs in Stassen's (2000) sample of 260 languages as an emphatic alternative to the prepositive monosyndetic coordination, such as in French:

French (Indo-European (Romance): France)

(264) a. Pierre **et** Sophia
PN and PN
'Pierre and Sophia'

b. **et** Pierre **et** Sophia
and PN and PN
'both Pierre and Sophia' (source: personal knowledge)

---

172. Lezgian also has the conjunction *wa* 'and', recently borrowed from Turkish:

šeher-r.i-n wa xür-er.i-n zehmetči-jar
town-PL-GEN and village-PL-GEN worker-PL
'the working people of towns and villages' (Haspelmath 1993: 330)

The usual strategy for coordinating NPs in French is the prepositive coordinator *et* 'and' (Example (264a)). In order to achieve an emphatic coordination ('both … and'), the coordinator is used twice and thus becomes a bisyndetic prepositive coordination (Example (264b)).

Postpositive bisyndetic coordination (*A(-)co B(-)co*) is common in the Caucasus, North-Eastern Africa, Southern India, Papua New Guinea and Australia (Stassen 2000). Dargwa is an example of a language with postpositive bisyndetic coordination:

Dargwa (Nakh-Daghestanian (Lak-Dargwa): Russia)

(265) dudeš.li-**ra** neš.li-**ra** emɦe b-abg-ili sa<b>i
father.ERG-and mother.ERG-and donkey.ABS N-harness-GER be:HPL
'Father and mother harnessed the donkey.' (van den Berg 2004: 199)

In (265) the conjunctive marker *-ra* 'and' suffixes to both units that are coordinated.

Mixed bisyndetic coordination is an extremely rare strategy (Haspelmath 2007: 10). Examples can be found in Latin and Homeric Greek:

Latin (Indo-European (Italic): present-day Italy)

(266) **et** singulis universis-**que**
and individuals universal-and
'both for individuals and for all together' (Haspelmath 2007: 11)

Homeric Greek (Indo-European (Greek): present-day Greece & E Mediterranean)

(267) Atreídēs **te** **kaì** Akhilles
PN's.son and and PN
'Atreu's son and Achilles' (Haspelmath 2007: 11)

In (266) we have a mixed bisyndetic coordination of the *co(-)A B(-)co* type. Notice that this is an emphatic coordination ('both… and') and that the unmarked conjunctive coordination for Latin is with the monosyndetic clitic *-que*. In (267) we have a mixed bisyndetic coordination of the *A(-)co co(-)B* type. Notice, furthermore, that in both cases the coordinators are two different markers.

11.1.2.3 *Multiple coordination*

In principle we may have an unlimited amount of units that get coordinated. It is common for languages to allow **coordinator omission** when more than two units are coordinated. In English, for example, coordinator omission is preferred: compare *John, Mary, Peter and Lucy all came to the party* with *John and Mary and Peter and Lucy all came to the party*. Retaining the coordinator between each unit makes for an emphatic kind of reading that is normally only used in special contexts. In English all but the last coordinator is omitted, as seems to be the most common strategy. However, there are languages where all but the first coordinator is omitted, such as Classical Tibetan.

Classical Tibetan (Sino-Tibetan (Bodic): present-day Tibet)

(268) a. sa-**dañ** tšhu-**dañ** me-**dañ** rluñ
earth-and water-and fire-and air
'earth and water and fire and air'

b. sa-**dañ** tšhu me rluñ
earth-and water fire air
'earth, water, fire and air' (Beyer 1992: 241)

In (268a) the coordinator *-dañ* 'and' is retained between each of the coordinated units, while in (268b) all but the first are omitted.

Not all languages allow coordinator omission. Sonsorol-Tobi, for example, requires a coordinator between each of the conjoined units:

Sonsorol-Tobi (Austronesian (Oceanic): Palau)

(269) Tekina **ma** Nninuo **ma** Yosio **ma** Kino le kae
PN and PN and PN and PN they often
pwi-wow dei-nako
go-DIRC fish-DIRC
'Tekina, Nninuo, Yosio, Kino often went out fishing.' (Oda 1977: 59)

In (269) the coordinator *ma* 'and' has to be placed between each of the conjoined NPs; something like **Tekina, Nninuo, Yosio ma Kino*, with all but one of the coordinators omitted, would not be acceptable in Sonsorol-Tobi (Oda 1977: 59).

11.1.3 Ellipsis

When clauses are coordinated this often means that some material is repeated, as in *Mary baked the cake and John ate the cake* (where the NP [*the cake*] is repeated). Many languages allow **ellipsis** (also called **coordination reduction** or **clause reduction**) where all but one occurrence of the repeated material is deleted, as in *Mary baked _____ and John ate the cake* (where the first occurrence of the NP [*the cake*] is deleted).[173]

Ellipsis can be either forward or backward. **Forward ellipsis**, or **analipsis**, is when the second of the coordinated units gets deleted:

(270) Mary bought a dress and John ____ a hat.

In (270) what would have been the second occurrence of the verb *bought* is deleted. This is also often called **gapping**, since in many European languages the most

173. In English this represents rather formal language. The informal equivalent would be something like *Mary baked the cake and John ate it*, where an anaphoric pronoun fills the slot of the deleted NP.

common form of analipsis is deletion of the repeated verb, which then leaves a gap between the preverbal and postverbal constituents (*John* ____ *a hat*).

Backward ellipsis, or **catalipsis**, is when the first occurrence of the repeated material gets deleted:

(271) Mary loves ____ and John hates romantic comedies.

In (271) what would have been the first instance of the repeated NP *romantic comedies* is deleted. Here it is the last element of the first coordinated clause that gets deleted (in this case the direct object *romantic comedies*); this is also called **right periphery ellipsis**.

11.2 Subordination

A **subordinated** (or **dependent**) **clause** is a clause that functions as a constituent, i.e. is **embedded**, within a **main** (or **superordinate**) **clause**. In *Mary said that she was going to take the bus*, for example, the full clause [*that she was going to take the bus*] functions as the object of the verb *said*. In other words, one of the arguments (in this case the object) is a full clause. Likewise a clause can function as an adjunct of a clause, as in *Mary will take the bus after she has finished packing her bag*, where the full clause [*after she has finished packing her bag*] functions as an adverbial. Full clauses may also function as modifiers of phrases, as in *Mary will pack the bag which John gave her for Christmas* where the clause [*which John gave her for Christmas*] functions as a modifier of the NP *the bag*, much as an adjective might (compare *Mary will pack the big bag* where the adjective *big* modifies the NP (*the*) *bag*).

The term **matrix clause** is sometimes used to denote a main clause. However it makes sense to distinguish between the two types of clauses. Following Quirk et al. (1985), I will here use the term **main clause** to denote an independent clause, i.e. a clause that is fully formed and functional on its own. Examples of main clauses are *Mary saw John*, *Mary told John that she saw him*, *Mary saw John as he got out of the bus*, *Mary saw the man who helped John get out of the bus*, and so on. I will, again following Quirk et al. (1985), use the term **matrix clause** to denote a clause **w**hich contains an embedded subordinate clause, i.e. where a subordinate clause functions as a constituent. That is, a matrix clause is a main (or superordinate) clause minus its subordinate clause part (Quirk et al. 1985: 991). Examples of matrix clauses are (underlined) *Mary told John that she saw him*, *Mary saw John as he got out of the bus*, and so on. In other words, a matrix clause is one which always carries a subordinate clause as one of its constituents. While a main clause is always independent, a matrix clause may or may not be independent. For instance, while *Mary saw John* functions on its own, **Mary told John* does not (since we lack the direct object of the verb *told*).

In this section I will define and give examples of complement clauses, adverbial clauses and relative clauses. The discussion will by necessity only scratch the surface of the complex issues related to subordinate clauses. For a very detailed analysis of subordination, including large cross-linguistic surveys, see Cristofaro (2005).

11.2.1 Complement clauses

A **complement clause** is a clause that functions as an argument of its matrix clause. This means that the matrix clause is incomplete without its embedded subordinate clause, since an argument constitutes an obligatory participant. A **subject complement** is a clause that functions as the subject of its matrix clause, as in *That the ball landed on his head startled the sleeping dog*, where the clause [*That the ball landed on his head*] functions as a subject to the verb *startled*. English usually **postposes** a subject complement and fills the subject slot with the pronoun *it*: *It startled the sleeping dog that the ball landed on his head*. An **object complement** is a clause that functions as the object of its matrix, such as *The sleeping dog dreamt that a ball landed on his head*, where the clause [*that a ball landed on his head*] functions as an object to the verb *dreamt*. A sentence can have several complement clauses embedded in each other:

(272) The dog hoped [that his master would realize [that it was time for dinner]].

In (272) the object complement [*that his master would realize*] in turn functions as a matrix for the object complement [*that it was time for dinner*].

Languages often have **complementizers**, that is, overt markers to signal that an entity is a complement clause. This may also be thought of as a subordinating particle, since the complementizer is flagging that the clause is subordinated to a matrix. The English complementizer is *that*, as in *The dog wished that he would get his food soon*, where *that* signals that the clause [*he would get his food soon*] is subordinated to the matrix *The dog wished*____.[174]

Complement clauses can be finite or non-finite. The former are more independent, or sentence-like (sometimes termed s-like) in that the complement clause looks similar in form to what it would have as an independent clause without the complementizer. The latter, non-finite complements, are less independent and do not look similar in form to a main clause.

The following will give some examples of different types of finite and non-finite complement clauses. It is by no means a comprehensive discussion of complementation. For a detailed and accessible overview of complementation from a typological perspective, see Noonan (2007), which this section relies heavily on. Dixon &

174. But see Noonan (2007) where *if* and *to* are also considered complementizers, as in *I don't know if he will come tomorrow* and *Jane wants Peter to invite her out*.

Aikhenvald (2006) gives not only very accessible general information on complementation, but also presents a collection of studies on complement clauses in different languages.

11.2.1.1 *Finite complements*

With **finite complements** the verb is typically finite in that it carries its own TMA inflection (and whatever necessary agreement), the subject is typically directly expressed (i.e. need not be the same as the one in the matrix clause), and the case marking of the arguments will typically be the same as in a main clause.[175] In short, the complement clause is **sentence-like** (s-**like**) in that it looks very similar in form to what it would have if it had been a main clause. An example of an English finite complement would be *Robert said that the new teacher sings very well*, where the complement clause [*that the new teacher sings very well*] could function as an independent (or main) clause without its complementizer *that*. The form of the elements in a complement clause may be similar to or the same as the form of the elements in a main clause, even though the constituent order changes. In German, for example, complements may be s-like morphologically but the word order for the subordinate clause goes from SVO to SOV:

German (Indo-European (Germanic): Germany)

(273) a. Der Mann sieht einen Film
ART.M.NOM man.NOM watch.3SG.PRES ART.M.ACC film
S V O
'The man is watching a film.'

b. Ich glaube,
1SG think.1SG.PRES
matrix clause
dass der Mann einen Film sieht
COMPL ART.M.NOM man ART.M.ACC film watch.3SG.PRES
S O V
complement clause
'I think that the man is watching a film.' (source: personal knowledge)

The complement clause of (273b) is s-like in that the form of the verb as well as of the arguments is the same as it would have been in a main clause (as in (273a)). The difference is only that the complementizer *dass* 'that' governs a verb final word order.

Some languages do not have any complementizer to flag that a clause is a complement clause, but the subordinate clause is simply juxtaposed. This is also called

175. But see Dixon (2006b: 45) for a discussion on why *finite* is a term that should be avoided.

parataxis or **apposition**. In Mualang, for example, the complement clause is finite and only prosody sets it off as a separate unit from the matrix clause:

Mualang (Austronesian (Malayic): Indonesia)

(274) ku N-dinga [ia udah datay]
1SG ACTV-hear 3SG PFCT come
S V O
matrix clause [complement clause]
'I heard [(that) he has come].' (Tjia 2007: 274)

In (274) the object of the matrix clause is the entire complement clause *ia udah datay* 'he has come'. There is no complementizer to flag the complement clause; only intonation sets the complement clause off as a subordinate to the matrix *ku Ndinga* 'I heard'.

11.2.1.2 *Non-finite complements*

Non-finite complements are less independent and typically are not formally similar to a main clause. The verb is typically not inflected for its own TMA and often the subject of the complement clause must be identical with that of the matrix clause. An example of a non-finite complement clause would be *Mary likes to play with dogs*, where the infinitival complement clause [*to play with dogs*] functions as an object to the verb *likes*. Notice that the complement clause could not function as an independent sentence (**To play with dogs* does not constitute a complete clause). Similarly, in *Mary enjoys playing with dogs*, the participial complement clause [*playing with dogs*] functions as an object and is incomplete as an independent clause (**Playing with dogs*). In both these example sentences the subject of the complement clause is the same as in the matrix clause: it is Mary who likes something and it is Mary who plays with dogs. Sentences like **Mary likes John to play with dogs* or **Mary enjoys John playing with dogs*, where the matrix and the non-finite complement clauses have different subjects, would not be acceptable.

In **infinitive complements** the verb of the complement clause is in its infinitive form and carries no syntactic relation to its notional subject (Noonan 2007: 67). In other words, there is no TMA or agreement marking on the verb. An example of an infinitive complement is *Mary tried to learn the poem*, where the verb of the complement clause [*to learn the poem*] is in its infinitival form (*to learn*). Even if the verb is in the infinitive and does not take any TMA or agreement marking, it still notionally is a verb (and not, for example, a nominalization). This can, for instance, be seen in the relation between the verb and the object in the complement clause. In English little more than the word order can tell us anything in that matter (but notice that the object occupies its usual slot, after the verb), but in other languages case marking on the object remains the same as it would have in a main clause:

Hungarian (Uralic (Ugric): Hungary)

(275) Anna meg-próbál-t-a meg-tanul-ni a vesset
PN PFX-try-PAST-DEF PFX-learn-INF the poem.ACC
'Anna tried to learn the poem.' (Kenesei et al. 1998: 33)

In (275) the infinitive complement (underlined) has its verb in the infinitive (marked by the suffix *-ni*) with no tense or agreement marker, as opposed to the verb in the matrix clause. The object in the complement clause is inflected for the accusative case, as it would have been in a main clause.

In **nominalized complements** the verb is made to function as a noun (it is nominalized), is the head noun of the NP and may take articles and case marking, may combine with adpositions and may in some cases be pluralized. An example of a nominalized complement clause is *The king's deer hunting gave entertainment to the whole court*, where the nominalized verb *deer hunting* functions as the head of the NP [*the king's deer hunting*], which in turn functions as the subject of the matrix clause.

Some languages allow various kinds of verbal marking on the nominalized verb. In Jamul Tiipay the nominalized verb takes the irrealis marker:

Jamul Tiipay (Hokan (Yuman): USA)

(276) [puu-ch w-yiw-x]-pu yaylly+pit
[that.one-S 3-come-IRR]-DEM forget
'I forgot (that) he was going to come.' (Miller 2001: 224)

In (276) the complement clause (in square brackets) functions as an NP, as evidenced by the demonstrative marker *-pu*, which attaches to the verbal noun *wyiwx-*. Even though the verb functions as a verbal noun, it takes the irrealis marker *-x*. The entire nominalized complement clause functions as the object of the matrix *yayl-lypit* '(I) forgot'.

Participial complements are somewhat different from other complement clauses in that syntactically they function as modifiers rather than as constituents. In other words, they function as adjectives or adverbs. An example of a participial complement is *I smelled Mary baking a cake*, with the verb in its participial form (*baking*). Here *Mary* functions as the object of the matrix and as the subject of the complement clause. The participle *baking* (*a cake*) functions as an adjective to the NP *Mary*. The modifying nature of the complement is, however, only syntactic and not semantic; it is the whole event and not only the argument that functions as the object of the matrix (*Mary*) that is being referred to. In other words, the sentence above does not mean 'I smelled Mary'.

11.2.2 Adverbial clauses

Adverbial clauses function as adjuncts to their matrix clauses. That is, they function as non-obligatory supplementary modifications to the verb phrase or the entire matrix clause. The term 'adverbial clause' thus refers to their "adverbial" function. Examples of adverbial clauses are *Mary saw John when he stepped out of the house* (time), *Mary lives where the road ends* (location), *Fold this as I showed you* (manner), *Mary phoned John because she missed him* (reason), and so on.

The following will give some examples of the kinds of adverbial clauses that can be found cross-linguistically. It is beyond the scope of this section to give a comprehensive discussion on adverbial clauses as a linguistic phenomenon. This section relies mainly on Thompson et al. (2007), which is a very accessible overview of adverbial clauses cross-linguistically. But see also Kortmann (1997) for a typology of adverbial clauses in the languages of Europe, including a discussion on the grammaticalization of adverbial clause subordinators in English, and van der Auwera (1998) for a collection of chapters giving various perspectives on adverbial constructions, predominantly on the languages of Europe.

Cross-linguistically adverbial clauses tend to fall into one of two groups: those that can be replaced with a single word and those that cannot (Thompson et al. 2007: 243). An example of an adverbial clause that can be replaced with a single word is *Mary saw John at the bus stop*, where [*at the bus stop*] may be replaced with, for example, *there* (or any other place adverbial) to form the sentence *Mary saw John there*. An adverbial clause of reason, such as *because she missed him* in *Mary phoned John because she missed him*, is not possible to replace with a single word adverb.

It should be noted that the types of adverbial clauses listed in the following are not claimed to be universal: a language may or may not use subordination to achieve the function of a given adverbial clause.

11.2.2.1 *Adverbial clauses that can be replaced with one word*

It is very common for languages to have specific adverbs expressing time, location and manner. Therefore, adverbial clauses that function as time, location or manner adverbials can generally be replaced with a single word. **Time clauses** function as a time adverbial and give information about the temporal sequence of events. This may either be indicated by a subordinating word, as in English (for example *John felt better **after** he had taken a nap*), or with a verbal affix. In Kamula the temporal adverbial clause is marked with a verbal sequence affix:

Kamula (Trans-New Guinea (Kamula): Papua New Guinea)

(277) koa yu solo-**po**, me-tle heta-wa
canoe water throw-IMM.SEQ there-ABL go.up-FP
'After we threw the water out of the canoe, we went up(stream).' (Routamaa 1994: 83)

In (277) the verbal affix *-po* marks sequentiality and indicates that the clause serves as a temporal adverb to its matrix.

Locative clauses function as place adverbs and are often introduced by a subordinator meaning 'where', as in *Mary sat **where** she always used to sit*.

Manner clauses function as manner adverbs and may either be indicated by a subordinate marker, as in *Peter felt **like** he could fly* or may take the form of a relative clause (see below), as in *Do it the way (that) I told you*.

11.2.2.2 *Adverbial clauses that cannot be replaced with one word*

While it is very common for languages to have specific adverbs expressing time, location and manner, it is rarely the case that languages have a specific adverb expressing purpose, circumstance, simultaneity, substitution, and so on. Therefore, the adverbial clauses discussed here can generally not be replaced with a single word.

A **purpose clause** expresses the purpose for an action. A **reason** (or **causal**) **clause** explains why an action or event took place. The difference between purpose and reason clauses is that "purpose clauses express a motivating event which must be *unrealized* at the time of the main event, while reason clauses express a motivating event which may be *realized* at the time of the main event" (Thompson et al. 2007: 251, emphasis in original). Consider the difference between the following sentences:

(278) a. He went out to get groceries.
b. He had a snack because he was hungry.

In (278a) the adverbial clause [*to get groceries*] explains why he went out, but the groceries had not been bought at the time he went out. The purpose for the action is thus still unrealized at the time of the main event. In (278b) the adverbial clause [*because he was hungry*] explains why he had a snack. His hunger was a real fact at the time he carried out the main event (had a snack). The explanation for the main event is thus a realized motivating event. Th unrealized nature of purpose clauses as in (278a) is often captured in the verbal morphology of the subordinate clauses. In Alacatlatzala Mixtec, for example, the subordinate verb in a purpose clause must be in the potential:

Alacatlatzala Mixtec (Oto-Manguean (Mixtecan): Mexico)

(279) a. ñā kihví kísā ún
it.INAN stupid CONT.do you.SG
chī sini ún kúmánī
because head your.SG CONT.lack
'You do stupid things because you lack sense.'
b. vashī rā **ñā** chīndēé rā ndihi
CONT.come he COMP POT.help he us.EXCL
'He is coming (in order) to help us.' (Zylstra 1991: 142, 146)

In (279a) we have a reason clause marked with the subordinator *chī*. The verb in the subordinate clause (*kúmanī* 'lack.CONTinuative') is inflected in the continuative, as is the verb in the matrix clause. In (279b) we have a purpose clause, indicated by the complementizer *ñā* and the fact that the verb in the subordinate clause is inflected in the potential (*chīndēé* 'help.POTential').

Circumstantial clauses express "the circumstances by which a given state of affairs comes to be" (Thompson et al. 2007: 253) and seem to represent a typologically rare type of adverbial clause (Payne 1997). An example of a circumstantial clause is *He got to the top at a young age by cheating in his work*, where the adverbial clause [*by cheating in his work*] states the circumstances that led to a situation described in the matrix.

A **simultaneous clause** indicates that two events overlap, as in *While he was jogging, he got a bright idea*, where the two events of jogging and getting an idea overlap. Languages may have a specific verbal marker to denote simultaneity:

Mauwake (Trans-New Guinea (Madang): Papua New Guinea)

(280) nefa ikum-**am**-ika-iwkin nan kerer-e-n
2SG.ACC illicitly-SS.SIM-be-2/3PL.DS there appear-PA-2SG
'They were just speculating about you when you arrived.' (Berghäll 2010: 121)

In (280) the same-subject simultaneous verbal marker *-am-* indicates that the two events of the speculation and the appearing overlap.

Conditional clauses can be divided into two basic semantic types, those that denote real conditions (**reality conditional**) and those that denote unreal conditions (**unreality conditionals**). Examples of reality conditionals are *If that milk is boiling, (then) you've kept it on the stove for too long* (present situation), *If you set an alarm clock, (then) it rings at the given time* (habitual/generic), *If you saw the film, then you know how it ends* (past). Unreality conditionals may be further divided into predictive situations (such as *If he comes today, (then) we'll have a party*) and imaginative situations. Imaginative situations may be of two kinds, hypothetical (*If I were rich, (then) I'd build a castle*) and counterfactual (*If you had read the news, (then) you'd know what I'm referring to*). The matrix of conditional clauses (the "then clause"), is sometimes called the **apodosis**, while the subordinate, the condition itself (the "if clause"), is sometimes called the **protasis**.

Concessive clauses make a contrast between the matrix and the adverbial clause. Examples of concessive clauses are *Although she went to bed early, she didn't feel rested the next day* (definite concessive clause) and *Wherever she went, she felt uncomfortable* (indefinite concessive clause). The difference between the two is that indefinite concessive clauses convey a tone of 'no matter what'.

Substitutive clauses denote that an expected event is replaced with an unexpected one, as in *Henry had dinner at a restaurant instead of going home*, where the adverbial clause [*instead of going home*] denotes the expected event that was substituted for the unexpected one.

Additive clauses express that one thing happened in addition to another, as in *Besides being late for their date, he was rude to her*, where the adverbial clause [*Besides being late for their date*] expresses what happened in addition to the event or state of affairs expressed in the matrix.

Those adverbial clauses that basically just serve to give a general background for the situation expressed in the matrix clause are sometimes called **absolutive clauses** (Thompson et al. 2007, Payne 1997) or **participial adverbial clauses** (Givón 2001b). The latter term reflects the fact that the verbs in absolutive clauses tend to be in a participial form in English, as in *Having finished the whole pot of honey, Winnie-the-Pooh contentedly licked his paws* where the adverbial clause [*having finished the whole pot of honey*] serves as a general background for the situation of Pooh licking his paws. Notice that the verb is in a non-finite form. Because the main function of this non-finite form is to indicate that the clause is a subordinated adverbial, the verbs in such clauses are sometimes called **converbs** (see, for example, Haspelmath 1995). Some languages use a general subordinator to mark an absolutive clause:

Godié (Niger-Congo (Kru): Ivory Coast)

(281) ɔ yi mɔ Dakpaduu´ **nʌ** bgesi ɔ tla a
he came to Dakpadu SUBORD traps he set recent
'Having come to Dakpadu, he set some traps.'
(Thompson et al. 2007: 266 citing Marchese 1976)

In (281) the subordinator *nʌ* marks the clause as subordinate, with the function of providing the background for the event in the matrix (the setting of traps). Notice that the verb here is not in any special (non-finite) form.

11.2.3 Relative clauses

Relative clauses are essentially clauses that modify an NP, and are sometimes also referred to as **adjective clauses**. A more specific definition for relative clauses is that they refer to clauses "narrowing the potential reference of a referring expression by restricting the reference to those referents of which a particular proposition is true" (Comrie & Kuteva 2011c). For instance, in the sentence *I met the actor who had performed the day before* the matrix has the referring expression [*the actor*]. The relative clause [*who had performed the day before*] narrows the potential referents to only the specific actor for which the proposition in the relative clause holds true, i.e. only that actor who had performed the day before.

In English a distinction is usually made between **restrictive** (or **defining**) **relative clauses** and **non-restrictive** (or **non-defining**) **relative clauses**. A restrictive relative clause is one that conforms to the above definition, namely a clause which restricts the potential referents in the matrix to those for which the proposition in the relative clause holds true. Compare *The man who sported a chequered felt hat spoke to me* with *The man, who sported a chequered felt hat, spoke to me*. The former sentence contains a restrictive relative clause which indicates that out of a number of men (for instance one wearing a cap, one wearing a black hat, one wearing a chequered felt hat, and one not wearing any hat at all), it was the one wearing a chequered felt hat that greeted me. Here the implication is that a sentence like *The man spoke to me* would not provide enough information to the hearer, making the added information [*who sported a chequered felt hat*] necessary in order for the hearer to identify precisely which man is being referred to. In the latter sentence the relative clause is non-restrictive and purely descriptive; it only adds some information about the man, but does not involve specifying one man out of several.

There are various formal differences between restrictive and non-restrictive relative clauses in English. One is prosody: non-restrictive relative clauses are usually set off from the matrix clause with pauses (indicated by commas in written text). Another is the choice of marker: non-restrictive clauses require relative pronouns, restrictive relative clauses also allow the relativizer *that* (compare *The man that sported a chequered felt hat spoke to me* with **The man, that sported a chequered felt hat, spoke to me*) or a gap (compare, for example, *The man* ____ *I spoke to yesterday came to my house* with **The man,* ____ *I spoke to yesterday, came to my house*). While there are other languages with formal distinctions between restrictive and non-restrictive relative clauses, this is very rare. For most languages it is therefore irrelevant to make a differentiation between the two types (cf. Comrie 1989:139). Because of this, the remainder of this section will deal with so-called restrictive relative clauses only, i.e. such clauses that conform to the definition given above.

Much has been written on relative clauses in general as well as the relative clauses of particular languages. For very accessible overviews of the typology of relative clauses, see Comrie (1989:138ff), Givón (2001b:175ff) and Andrews (2007a). A very thorough study indeed on the typology of relative clauses is Lehmann (1984), unfortunately accessible only to those who read German.

11.2.3.1 *The structure of the relative clause*

There are several pertinent parts to a sentence that contains a relative clause, each of which I will briefly mention. Consider the following sentences:

(282) a. I know *the man* [*who waved at us*]
b. *The man* [*that waved at us*] seems familiar to me

The matrix clause is the clause in which the relative clause is embedded (*I know the man* and *The man seems familiar to me* respectively). The **head** of the relative clause is the NP that the relative clause modifies (*the man* in both sentences). The head of the relative clause may also be termed $\mathbf{NP}_{mat}$ since it is an NP in the matrix clause that is being delimited by the relative clause. The **restricting clause**, i.e. the clause that delimits the reference of the referring expression, is the relative clause itself. This may also be termed $\mathbf{S}_{rel}$. The S_{rel} will, in this section, be indicated with square brackets (as in Example (282)). A relative clause may be flagged, or set off, by a **relativizer**, as in (282b) *that* and an S_{rel} may or may not contain an overt element that is coreferential with the head (or NP_{mat}). In (282a) there is such an overt element, a **relativized NP** (or $\mathbf{NP}_{rel}$): *who* is an NP_{rel} which is coreferential with the head NP (or NP_{mat}) *the man*. In (282b) there is no such coreferential element; instead there is a gap where the NP_{rel} would have been: ***The man that*** [____ ***waved at us***] *seems familiar to me*.

It is important to notice that the head (NP_{mat}) and the relativized NP (NP_{rel}) do not necessarily have to have the same syntactic roles. Compare the following:

(283) a. I know *the man* [*who waved at us*]
S V O [S V O]
b. *The man* [*who(m) I waved to*] recognized me
S [O S V] V O

In (283a) the NP_{rel} (*who*) is the subject of the relative clause but the NP_{mat} is the object of the matrix clause. In (283b) the NP_{rel} (*who(m)*) is the object of the relative clause but is coreferential with the subject of the matrix.[176]

Languages have different strategies for representing the NP_{rel}. We have seen that English may leave a gap for the NP_{rel}. The **gap** strategy is extremely common, spread all over the world. Of the 166 languages in Comrie & Kuteva's (2011b) database on the relativization on NP_{rel} subjects 125 (or 75.3%) have a gap. Notice that this refers to any strategy where there is no overt element that is coreferential with the head NP. Whether a relativizer is present or not is irrelevant:

Turkish (Altaic (Turkic): Turkey)

(284) [____ balığ-ı yi-yen] adam
fish-ACC eat-PTCPL man
'the man [who eats/ate the fish]' (Kornfilt 2003: 58)

---

176. The S_{rel} [*who(m) I waved to*] exhibits the typologically rare phenomenon of **preposition stranding**, where the preposition is left hanging at the end of the clause because the NP_{rel} has been fronted to clause initial position (compare *I waved to the man*, where the NP *the man* follows the preposition *to*). It is more common for the prepositional phrase to have its conventional order of preposition + NP, which is also possible in English, though it is more formal (*the man* [*to whom I waved*]). The latter strategy is informally termed **pied-piping**. The complexities of preposition stranding versus pied-piping will not be discussed further here.

In (284) there is no relativizer equivalent to the English *that*. The strategy is still a gap strategy, since there is no overt reference to the head NP (*adam* 'man') in the S_{rel}. Notice also that even though the verb is in a non-finite form, the clause follows the definition of a relative clause in that it serves the function of narrowing the potential referents of the head NP.

English may also have a **relative pronoun** (*who* or *whom* depending on syntactic role), i.e. a pronoun inside the S_{rel} that is coreferential with the NP_{mat} and that at the same time functions as a relativizer. This strategy is found in only 12 languages (or 7.2%) for subject NP_{rel}s, all of them, except Acoma, European languages.[177] A somewhat more common strategy for subject NP_{rel}s is **non-reduction**, where the head noun (re)appears as a full NP in the S_{rel}, something along the lines of *The man* [*I waved to the man*] *recognized me*. This is found in 24 (or 14.5%) of the languages in Comrie and Kuteva's database, most of them in the Americas. The strategy non-reduction actually covers three different subtypes in Comrie & Kuteva's database: internally headed clauses, correlative clauses and paratactic relative clauses. In **internally headed relative clauses** (also called **internal relative clauses**) the head NP (NP_{mat}) appears inside the relative clause and is not represented in the matrix. Jamul Tiipay is an example of a language with such a strategy:

Jamul Tiipay (Hokan (Yuman): USA)

(285) ['iipa peya nye-kwe-'iny]-pe-ch mespa
[man this 3/1-S.REL-give]-DEM-SUBJ die
'The man [who gave me this] died.' (Miller 2001: 207)

In (285) the head NP *'iipa* 'man' appears not in the matrix but in the S_{rel} itself.

In **correlative clauses** the head NP also occurs inside the S_{rel}, but as opposed to internally headed relative clauses, the head NP in correlative clauses has an overt reference in the matrix (either anaphorically with a pronoun or with a repetition of the head NP). Supyire is an example of a language with correlative clauses:

Supyire (Niger-Congo (Gur): Mali)

(286) [pùcwòyì u ɲyɛ ná mu í ke]
[girl.DEF(I.SG) she(I.SG) be with you with REL]
uru sí ŋ̀-kwû
she(EMPH.I.SG) FUT FUT-die
'The girl [who is with you] will die.' (lit. [The girl she who is with you] she will die)
(Carlson 1994: 489)

177. English is listed in this group in Comrie & Kuteva's (2011b) database, although it also has other strategies.

In (286) the head NP *pùcwòyì* 'girl' appears inside the S_{rel} and is anaphorically connected with the pronoun *uru* 'she' in the matrix. The relativizer *ke* marks the clause as an S_{rel}.

A **paratactic relative clause** "contains the full-fledged head and is the same as an unmarked simple (declarative) clause; the relative and main clauses are only very loosely joined together" (Comrie & Kuteva 2011b). Amele is an example of a language with paratactic relative clauses:

Amele (Trans-New Guinea (Madang): Papua New Guinea)

(287) mel mala heje on
boy chicken illicit take.3SG.SUBJ.REM.PST
((mel) eu) busali nu-i-a
boy that run.away go-3SG.SUBJ-TOD.PST
'The boy [that stole the chicken ran away].'
(Comrie & Kuteva 2011b citing John Roberts p.c.)

In (287) the head NP *mel* 'boy' appears in what might be termed the relative clause (even though it is not overtly subordinate). It may optionally be repeated in what might be termed the matrix, together with an optional determiner. Intonation indicates that these are not two independent main clauses. Notice also that, unlike in the correlative clause above, there is no relativizer of any kind.

The last strategy coded in Comrie & Kuteva (2011b) is **pronoun retention**. This is very rare, found only in five languages (3%) in their sample: Babungo, Baka (Niger-Congo (Ubangi): Cameroon), Eastern Kayah Li, Ngemba (Niger-Congo (Bantoid): Cameroon) and Yoruba.

Babungo (Niger-Congo (Bantoid): Cameroon)

(288) wə [ntɨə fáŋ ŋwə nə sàŋ ghɔ̀] jwì féenə
person [that who he PAST4 beat.PF you] come.PF here
'The man [who has beaten you] has come here.' (Schaub 1985: 33)

In (288) the relative clause contains a resumptive personal pronoun (*ŋwə̀* 'he') which is coreferential with the head NP.

A strategy not included in Comrie & Kuteva's map is that of **headless relative clauses** (also called **free relative clauses**), where the head NP is lacking altogether, such as [*Whoever disturbs me now*] *will get a scolding*, where the reference to the NP_{mat} is in the S_{rel} itself (cf. *Any person* [*who disturbs me now*] *will get a scolding*).

Languages do not necessarily use the same relativization strategy for different syntactic roles. In English, for example, the gap strategy is not acceptable with possessors (compare *I love the dog whose ears are floppy* with **I love the dog that* ____ *ears are floppy*). The pattern described above for relativization of NP_{rel} subjects thus differs somewhat from the pattern in Comrie & Kuteva's (2011a) database on relativization of NP_{rel} obliques, as shown in Table 11.1:

Table 11.1 Comparison between relativization strategies for subjects and obliques. Adapted from Comrie & Kuteva (2011a and 2011b). Absolute numbers in parentheses.

	Value	Subjects		Obliques	
1.	Relative pronoun	7.2%	(12)	11.6%	(13)
2.	Non-reduction	14.5%	(24)	12.5%	(14)
3.	Pronoun retention	3%	(5)	17.9%	(20)
4.	Gap	75.3%	(125)	49.1%	(55)
	Total		166		112[178]

The differences in these patterns conform to the Accessibility Hierarchy of Relativization formulated by Comrie & Keenan (1977). The hierarchy captures the general cross-linguistic pattern that subjects are more likely to be relativized than objects, which are in turn more likely to be relativized than possessives.[179] Put differently, the further to the left on the hierarchy, the easier it is for that position to be relativized.

subject > direct object > indirect object > oblique > possessor

Figure 11.1 The Accessibility Hierarchy of Relativization.

The hierarchy postulates that subjects are the easiest to relativize. It further postulates that if a language allows relativization on direct objects, it will also allow relativization on subjects, and if a language allows relativization on indirect objects, it will also allow relativization on direct objects and subjects, and so on. Moreover, it generalizes that explicitness increases further down the hierarchy, so that the least explicit (or overt) strategy is more likely to be found further to the left and the more explicit (or overt) strategies are likely to be found further to the right. This is borne out by Comrie & Kuteva's databases, since the gap strategy is more common with subjects than with obliques, while pronoun retention, which is a more explicit strategy, is more common with obliques than with subjects.

11.2.3.2 *The position of the relative clause*

We have seen that relative clauses can be either internally headed or externally headed. It is vastly more common to have externally headed relative clauses than internally headed ones. Of the 825 languages in Dryer's (2011p) database on the order of the

178. There is a fifth value 'not possible' in Comrie & Kuteva (2011a) containing ten languages (8.9% of the sample).

179. This of course presupposes that all languages have identifiable categories such as 'subject' and '(direct/indirect) object'. For a discussion on why those categories are not universal, see, for example, Andrews (2007b) and LaPolla & Poa (2006).

relative clause and head NP, only 24 (or 2.9%) are internally headed. Eight languages (1%) have adjoined relative clauses while seven languages (0.8%) have correlative relative clauses and only one language (0.1%), Kombai (Trans-New Guinea (Awju-Dumut): Indonesia), is listed as having 'double-headed' relative clauses.[180] In contrast, a full 721 languages (or 87.4%) have externally headed relative clauses. The remaining 64 languages (7.8%) have a mixed system.

The externally headed relative clauses can be divided into two groups, those where the relative clause follows the head NP (**postnominal relative clauses**, or **NRel** for the order (head) noun-relative clause), and those where the relative clause precedes the head NP (**prenominal relative clauses** or **RelN** for the order relative clause-(head) noun). The former type, NRel, is by far the most common type, with 580 languages (80.4% of the languages with externally headed relative clauses or 70.3% of the entire sample), while RelN is found in 141 languages (19.6% of the languages with externally headed relative clauses or 17.1% of the entire sample) in Dryer's database. English is an example of a language with a predominantly NRel order (*The man*$_{N}$ [*who wore a felt hat*]$_{Rel}$) while Evenki is an example of a language with RelN as its dominant strategy:

Evenki (Altaic (Tungusic): Russia)

(289) bi [tatkit-tu haval-d'a-cha-ve]$_{Rel}$ beje-ve$_{N}$ archa-Ø-m
I [school-DAT work-IPFV-PTCPL-ACCD] man-ACCD meet-NFUT-1SG
'I met the man [who worked at school].' (Nedjalkov 1997: 36)

In (289) the S_{rel} precedes the head NP *beje* 'man'. Turkish is another example of a language with RelN as its dominant strategy; cf. Example (284). Languages with NRel order are spread all over the world, while languages with RelN as their dominant strategy are found predominantly in Asia (except for Southeast Asia and the Middle East).

Taking out the contact languages (Ndyuka and Sango, both with externally headed relative clauses and NRel order) does not affect the pattern for the WALS languages noticeably. While the figures for the languages in APiCS are at the time of writing still temporary for this feature, it seems that the pattern for the order of the relative clause and noun differs somewhat from that in WALS (Michaelis et al. 2013: feature 7). The absolute majority of the sample has NRel as the dominant strategy while only very few

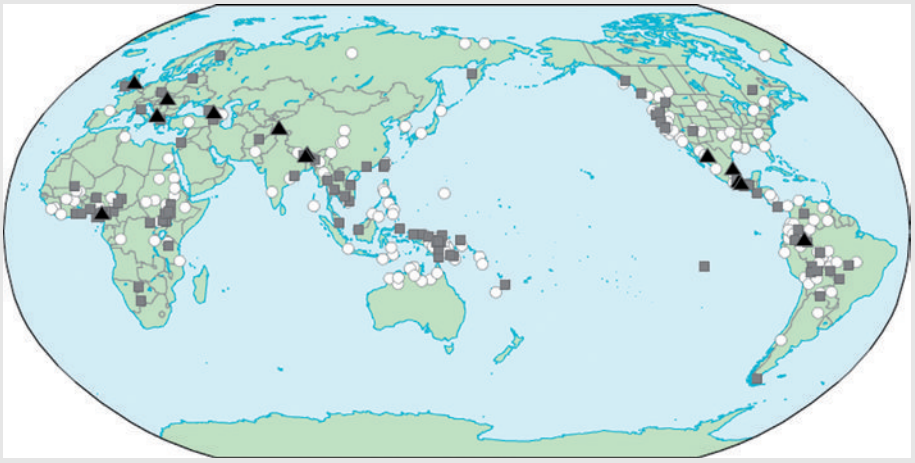
180. What Dryer terms '**double headed relative clauses**' is essentially the same thing as correlatives except that the full head NP is repeated in the relative clause. In **adjoined relative clauses** the S_{rel} appears outside the NP_{mat} and may be separate from it. The head NP is in the matrix and is not anaphorically referred to in the S_{rel}:

Martuthunira (Australian (Pama-Nyungan): Australia)
ngayu nhawu-ngu-layi kanyara-lu [jalyuru-ma-rnura-lu]
1SG.NOM see-PASS-FUT man-FUT [hole-CAUS-PRES.REL-EFF]
'I'll be seen by the man [who is digging a hole].' (Dench 1995: 244)

languages seem to have RelN as their dominant strategy. In other words, externally headed relative clauses seem even more common for the languages in the APiCS than for those in WALS.

ANIMACY DEPENDENT RELATIVE CLAUSE MARKERS

In English the choice of relative pronoun *who*(*m*)/*which* agrees with the animacy of the head NP, as in *The man who greeted me* (but not **The man which greeted me*) and *The stone which fell* (but not **The stone who fell*). This is quite rare cross-linguistically. In a survey of 275 languages I found that 108 had some kind of relative clause marker (grey squares). Of these, the relative clause marker was animacy dependent in only 13 languages (black triangles). White dots on the map indicate languages that form relative clauses by other means than with a relative clause marker. For a full legend, see http://dx.doi.org/10.1075/z.176.additional.



The English animacy dependency is not inherited: other Germanic languages either have an invariant marker or have gender dependent relative pronouns. Swedish, for example, has an invariant marker, while German has gender dependent relative pronouns:

Swedish (Indo-European (Germanic): Sweden)
flickan [som vinka-de]
girl.DEF [REL wave-PAST]
'The girl [who waved]'
stenen [som föll]
stone.DEF [REL fall.PAST]
'The stone [which fell].' (source: own knowledge)

German (Indo-European (Germanic): Germany)
das Mädchen [das wink-te]
ART.N.NOM girl [RELP.N.NOM wave-PAST]
'The girl [who waved]'
der Stein [der fiel]
ART.M.NOM stone [RELP.M.NOM fall.PAST]
'The stone [which fell].' (source: own knowledge)

In Swedish the relative marker is the same irrespective of the semantics of the head NP. In German the relative pronoun agrees with the head NP in gender and number. Notice that the gender is not animacy dependent: *Mädchen* 'girl' (human) is in the neuter while *Stein* 'stone' (inanimate) is in the masculine gender.

The order of the relative clause and noun forms a unidirectional correlation with the order of object and verb. Of the 879 languages in Dryer's (2011z) database on the relationship between the order of the object and verb and the order of the relative clause and noun, 245 languages have OV word order. Of these, 132 (53.9% of the OV languages) have the order RelN, while 113 (46.1% of the OV languages) have the order NRel. There is no significant correlation here. However, of the 421 languages with the order VO a full 370 (87.8% of the VO languages) have the order NRel while only five (1.2% of the VO languages), Amis (Austronesian (Paiwanic): Taiwan), Bai, Cantonese, Hakka and Mandarin, have the order RelN. We may thus make the unidirectional implicational statement that if a language is VO, then it is likely to be NRel.[181]

11.3 Cosubordination

The term **cosubordination** was originally coined by Michael Olson (1981) to describe a type of complex clause in Barai (Trans-New-Guinea (Koiarian): Papua New Guinea) that had similarities with both coordination and subordination. Since then the term has mostly been used in the Role and Reference Grammar approach of linguistic analysis (cf. Foley & Van Valin 1984 and, for example, Van Valin 2005, with references).[182] It is a convenient term that neatly captures the essence of the kinds of complex clauses that are not embedded (which makes them similar to coordinated clauses) but are also not independent (which makes them similar to subordinate clauses). The difference between coordinate, subordinate and cosubordinate clauses are illustrated in Figure 11.2 (from Van Valin 2005: 188):

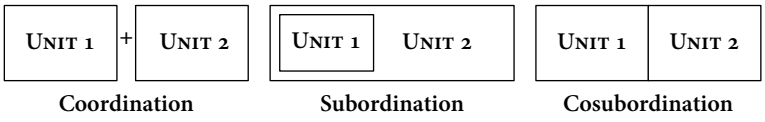


Figure 11.2 The difference between coordination, subordination and cosubordination. From Van Valin (2005: 188) with minimal changes.

181. The remaining 213 languages in Dryer's (2011z) database are listed as 'other'. These languages either have internally headed relative clauses or do not have any dominant word order.

182. Role and Reference Grammar (RRG) is an analytical framework that seeks to integrate both description and theory in its linguistic analysis. For a detailed and recent book length introduction to RRG, see Van Valin (2005). For a shorter overview of RRG and a continuously updated bibliography, as well as a generous list of downloadable works, see the RRG page: http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html (accessed 11 March 2011).

As Figure 11.2 illustrates, the essence of coordinated clauses is that no unit is embedded in any other and each may function independently as a main clause. This is indicated by the separate boxes for each of Unit 1 and Unit 2, which are then linked together with some kind of coordination (the "+" in the figure). With subordination, on the other hand, the subordinated clause is embedded in the matrix: Unit 1 in Figure 11.2 is embedded in Unit 2. With cosubordination, on the other hand, none of the units is embedded in the other. However, they are also not separate from each other; they are not two independent units, as indicated by the fusion of the two boxes.

This section will briefly define and give examples of serial verb constructions and clause chaining, two major types of cosubordination. Each of these constructions is highly complex and deserves much more space for discussion than this section is able to provide.

11.3.1 Serial verb constructions

A **serial verb construction** (also simply called **serial verbs**) is "a sequence of verbs which act together as a single predicate, without any overt marker of coordination, subordination, or syntactic dependency of any other sort. Serial verb constructions describe what is conceptualized as a single event" (Aikhenvald 2006c: 1). In other words, a series of two or more verbs, which are not compounded and do not belong to separate clauses, are stringed together to describe one event. English does not have serial verb constructions; the closest examples would be certain kinds of commands in colloquial English (especially in American English) such as *Run go fetch the ball!*, where [*run go fetch*] jointly serve as the predicate of the clause. Here the command is not that the addressee should first run, then go and then fetch something. The command is about fetching the ball, and the construction conveys a sense of both urgency and motion, something to the effect of "off you go, fetch the ball and do it now". Serial verb constructions are quite common cross-linguistically and are especially prominent in West Africa, Southeast Asia, Papua New Guinea, among the languages of Amazonia and among Oceanic languages. Roughly half of the languages in APiCS have serial verb constructions (Michaelis et al. 2013: features 84–86), indicating that they are also common for pidgins and creoles.

This section will give some examples of the most prototypical characteristics of serial verb constructions. I stress that these characteristics are meant only as *prototypical* characteristics. Individual languages will have various kinds of serial verb constructions where the given characteristics hold to a higher or lesser degree. For an accessible introduction to the typology of serial verb constructions, see Aikhenvald (2006c), which serves as the introduction to Aikhenvald & Dixon (2006b), where examples and discussions of serial verb constructions in a number of diverse languages can be found. A shorter and very accessible survey of serial verb constructions, including their developmental processes, is Bisang (2009).

Prototypical serial verb constructions refer to **one single event**. Consider the following example:

Loniu (Austronesian (Oceanic): Papua New Guinea)

(290) u **pin** **imon** **ime** amat
1DU R.NSG.change R3SG.return R3SG.come human
'We (two) changed back into human form.' (Hamel 1993: 124)

In (290) the serial verb construction denotes one single event, the changing back (into a certain shape); a translation like 'we changed (and then) we returned (and then) we came' would be incorrect.

Serial verb constructions typically form **a single clause**, which means that they do not take overt markers of syntactic dependency, such as coordinating or subordinating particles. In Logba, for example, a serial verb construction takes the person agreement marker only on the first verb and allows no overt linkers. The event denotes one single time frame, meaning that we have one event which is the sum of the semantics denoted in the verbs:

Logba (Niger-Congo (Kwa): Ghana)

(291) a. selorm **ó-huitɛ** **bá** a-fán
PN SM.SG-run come CM-house
'Selorm ran home.'

b. selorm **ó-huitɛ** yɛ ɔ-**bá** a-fán
PN SM.SG-run CONJ 3SG-come CM-house
'Selorm ran and (then) came home.' (Dorvlo 2008: 193)

In (291a) we have one single clause denoting the single action of Selorm running home. Notice that the verbal agreement marker occurs only once, on the first verb. In (291b) we have two coordinated clauses, as indicated by the coordinating conjunction *yɛ* 'and'. This construction conveys a meaning of sequentiality involving two separate actions, that of running and that of coming home: "[i]ndeed it gives the impression that the man engaged in a race and after that he came home" (Dorvlo 2008: 193). Notice that here the predicates of each clause get separate verb agreement markers, another indicator that the verbs in (291b) belong to one clause each and do not function as a single predicate in a single clause.

Serial verb constructions typically **share grammatical categories** such as TMA, negation or person agreement (as in Logba in (291) above) and typically **share at least one argument** in the clause. In some languages the TMA marker can only occur once for the entire construction, as in Example (292) below, while in other languages the TMA marking is repeated for each of the verbs of the construction, as in Example (290) above. Note, however, that in languages where the TMA marking is repeated, the individual verbs typically need to have the same TMA (i.e., to have some of the verbs in realis and some in irrealis mood in (290) above would not be acceptable).

Lokaa (Niger-Congo (Cross River): Niger)

(292) úbi **nà** **ọ́-kọ́bí** lò-wí **ó-gaŋa** ká gbángbán
PN FUT I.AGR-fetch XIII-water I.AGR-put in basin
'Ubi will fetch water and put it in the basin.' (Baker 2005: 133)

In (292) the future marker *nà* may only occur once as its scope covers both verbs of the construction (*ọ́kọ́bí* 'fetch' and *ógaŋa* 'put' respectively); something like **úbi* ***nà*** *ọ́kọ́bí lòwí* ***nà*** *ógaŋa ká gbángbán*, where the future marker is repeated for each verb, would not be acceptable.[183] We also see in Example (292) that the two verbs share arguments: *Úbi* is the subject of both verbs and *lòwi* 'water' is the object of both verbs.

Serial verbs typically behave as **one single predicate**, in that, for example, nominalizing or relativizing markers appear only once:

Toqabaqita (Austronesian (Oceanic): Solomon Islands)

(293) kwaqe fole-la-na niu qe aqi si qafetaqi
hit split-NZR-3PL coconut 3SG.NFUT NEG.V 3SG.NEG be.difficult
'Splitting coconuts is not difficult.' (Lichtenberk 2006: 261)

In (293) the nominalizing suffix appears only on the last of the verbs in the serial verb construction, but its scope covers the entire construction (as does the 3rd person marker), showing that the construction functions as one single predicate. The closest English equivalent would be something like "hit-splitting coconuts is not difficult". Furthermore, serial verbs typically form **one intonational unit**, that is, they do not have pauses between the individual verbs of the construction, which again makes them similar to monoverbal clauses (i.e. clauses with only one verb). In Toqabaqita, for example, the serial verb constructions behave like one unitary word and are uttered in one intonation contour (Lichtenberk 2006: 262).

11.3.2 Clause chaining

A second type of cosubordination is **clause chaining**, where two or more clauses are combined without any overt syntactic marker (i.e. no overt linker indicating coordination or subordination). With chained clauses there is typically only one in the string (or chain) that is inflected for TMA, while the other clauses in the chain are in a non-finite and usually more restricted form (e.g. participial). In other words, two or more clauses clauses are jointly subordinated to a matrix in a chain, the latter being the clause that carries the TMA information, "like an engine that pulls a string of cars" (Longacre 2007: 398). Most commonly, but by no means always, the matrix is the last

183. For a definition of scope, see 12.1.1.2.

clause in the chain. Due to the fact that the matrix, carrying the distinctive marking, is usually the last clause, this is often called the 'final clause' as opposed to the 'non-final clauses' coming before and carrying a more restricted marking. The term 'non-final clause' has largely been replaced with the term **medial clause**, to capture the fact that these clauses usually have a medial position in the string. Notice that the cosubordination here refers to the medial clauses, which are not embedded in each other, but also do not function independently.

Medial clauses and serial verbs differ in two very significant ways. First of all, contrary to serial verbs, chained clauses typically have independent marking to indicate whether the following clause has the same or a different subject. The medial verb (i.e. the verb of the medial clause) is thus typically marked twice, once for its own subject and once to indicate whether the following clause (or, in some cases, the final clause) has the same or a different subject, so-called **switch reference**. Secondly, chained clauses typically refer to a series of events, as opposed to serial verbs which refer to one single event. Thus, chained clauses also often express temporal relations with respect to the other clauses in the chain such as whether the events overlap or occur in succession to each other (Longacre 2007: 399). Clause chaining is typically used for discourse information structuring whereby long chains of medial clauses have a narrative relationship as, for example, a foregrounding or backgrounding device.

This section will give some examples of typical properties of medial verbs as well as the argument tracking device of switch reference. It is by necessity merely a brief overview of these very complex features. Longacre (2007) discusses clause chaining at length. For more on clause chaining and foregrounding/backgrounding of information, see, for example, Dooley (2010) with references. Much information on clause chaining can be found in Haiman & Thompson (1988). For a thorough discussion on switch reference in a formalized analytical framework, see Stirling (1993).

11.3.2.1 *Medial clauses*

Medial clauses, or **medial verbs**, are those clauses (or verbs) that are chained together cosubordinately to form a structural narrative completed by the finite clause (the matrix, sometimes also called the **reference clause** in clause chaining contexts). Clause chaining is common in Papua New Guinea, Australia and the Americas (especially northern South America and south-western USA), as well as in a belt from the Caucasus stretching through Central Asia all the way to Korea and Japan in Northeast Asia (Longacre 2007). Mauwake is an example of a language with clause chaining (medial verbs underlined):

Mauwake (Trans-New Guinea (Madang): Papua New Guinea)

(294) a. wien **en-emi**, epira lolom **if-emi** ne owowa
3PL.GEN eat-SS.SIM plate mud spread-SS.SIM ADD village
p-urup-em-ik-e-mik
BPX-ascend-SS.SIM-be-PAST-1/3PL
'They ate it themselves, spread mud on the plates and brought them up to the village.'

b. nomokowa maala **war-ep** **ekap-ep** ifa nain **ifakim-o-k**
tree long cut-SS.SEQ come-SS.SEQ snake that kill-PAST-3SG
'He cut a long stick, came and killed the snake.' (Berghäll 2010: 300, 304)

In (294a and b) only the last verb of the sentence carries the past tense marker (*-e-* and *-o-* respectively) while none of the chained medial verbs carry any TMA marking at all. Notice that the chain of events is coded for overlap versus succession: in (294a) the medial verbs are coded for overlap with the same subject simultaneous marker *-em(i)-*, while in (294b) the medial verbs are coded for succession with the sequential marker *-ep-*. Notice also that in both sentences the chained clauses refer to a series of events: (294a) refers to the eating, the spreading of mud and the bringing of the plates, three separate events, while (294b) refers to the cutting of a stick, the coming, and the killing of the snake, again three separate events.

While the majority of clause chaining constructions place the finite clause (matrix or reference clause) last in the sequence, there are also languages where it appears initially. In Logoti, for example, the finite (matrix or reference) clause appears sentence initially, with the chained medial clause following:

Logoti (Nilo-Saharan (Moru-Ma'di): DR Congo)

(295) yìko **dré** bhógò bí **a-kí-zó** líɲyá **bhà-zo**
brother.in.law R banana leaf DIRC-gather-SRF fufu put-SRF
àto **bhà-zo** **dò-zo-á** **adzí-lé** agó nda dré
vegetable put-SRF take-SRF-3OBJ send-ORF man that to
'The brother-in-law gathered banana leaves, added fufu, added vegetables, took these, and sent them to that man.' (Wright 1995: 106)

In (295) the first verb in the chain is the finite one, inflected with the realis and directional markers (*dré* and *a-* respectively). All other verbs form medial clauses, chained together and dependent on the initial matrix (or reference) clause. Notice, again, that the chained clauses refer to separate events: the gathering of leaves, the putting of *fufu*, the putting of vegetables, the taking of the items and the sending them off – five separate events.

11.3.2.2 *Switch reference*

Another typical feature of clause chaining is, as mentioned, **switch reference**, which is basically a referent tracking device indicating if the subject of the following clause is the same as or different from the one in the present clause. Consider the examples from Amele below:

Amele (Trans-New Guinea (Madang): Papua New Guinea)

(296) a. ija hu-m-ig sab j-ig-a
1SG come-SS-1SG food eat-1SG-TOD.PAST
'I came and ate the food.'

b. ija ho-co-min sab ja-g-a
1SG come-DS-1SG food eat-2SG-TOD.PAST
'I came and you ate the food.' (Roberts 1988: 49)

In (296a) the finite verb is marked for TMA (the suffix *-a* 'TODay's PAST') as well as for agreement (*-ig* '1st sinGular'). The medial verb *hu-* 'come' is marked with the same subject suffix *-m* to indicate that the subject of the verb is the same as for the following verb. In (296b) the switch reference marker changes to the different subject suffix *-co* to indicate that the subject of the verb is not the same as the following verb. Notice that switch reference is not necessarily the same as verbal agreement: both verbs in (296) are inflected for verbal agreement with their argument.

In some languages reference tracking is part of the verbal agreement. For these languages the third person may be divided into two categories, essentially coding "third person same subject" and "third person different subject". Central Yup'ik is an example of such a language:

Central Yup'ik (Eskimo-Aleut (Eskimo): USA)

(297) a. Dena-q quya-u-q Toni-aq cinga-llra-Ø-ku
PN-ABS happy-INTR-3.ABS PN-ABS greet-because-3/3-DEP
'Dena$_i$ is happy because she$_j$ greeted Toni.'

b. Dena-q quya-u-q Toni-aq cinga-llra-**mi**-ku
PN-ABS happy-INTR-3.ABS PN-ABS greet-because-SS-DEP
'Dena$_i$ is happy because she$_i$ greeted Toni.' (Payne 1997: 322)

In (297a) Dena is happy because somebody else greeted Toni, which is indicated with the verbal 3/3 agreement marker Ø (and noted in the translation with the subscript *i* for Dena and *j* for the third person who is not coreferential with Dena). In (297b) Dena is happy because she herself greeted Toni, indicated with the verbal agreement marker *-mi* which denotes third person same subject (and indicated in the translation with the subscript *i* showing that *Dena* and *she* refer to the same person). The terms for these kinds of systems may vary; 'third person same subject' marker is sometimes referred to as, for example, 'reflexive', 'fourth person' and 'recurrent'.

Many languages have portmanteau markers for reference tracking and temporal relation. In Eastern Pomo, for example, the reference tracking devices denote mood in that there are different markers for realis and irrealis clauses. Furthermore, in the realis mood two temporal references are distinguished, overlap or simultaneity ('coincident' events in Mithun's (1993) terminology) and succession or sequentiality ('consecutive' events in Mithun's terminology). This leads to six different markers:

Eastern Pomo (Hokan (Pomoan): USA)

		SAME SUBJECT	DIFFERENT SUBJECT
IRREALIS		hi	hla
REALIS	SIMULTANEOUS	in	da
	SEQUENTIAL	ba	li

(Mithun 1993: 121)

(298) a. qʰá=·l yó-**hi** maʔá qʰa·díway=ʔkʰe
water=to go-SS.IRR food buy=FUT
'He$_i$'ll go down and (he$_i$'ll) buy groceries.'

b. čhé múl-aq=**hla** ya čhó-w=ʔkhe hlí-w
rain fall-PL=DS.IRR 1PL not-PFV=FUT go.PL-PFV
'If it rains, we won't go.'

c. čá·č'=ʔel mu·l ba·-dáqʰa·-del qʰdé·ʔ-č'-du-**n**
man=DEF that POSS-wife-PAT fight-REFL-IPFV-SS.R.SIM
m-čó-t̲-**in** ma· pʰ-ts'áday
kicking-set-ME-SS.R.SIM ground swinging-slam
'The man, fighting with his wife, kicking her, slammed her to the ground.'

d. má·t̲a=ʔel maʔá yhé-·n=**da** báya·=ʔel k'úč-i·
woman=DEF food do-IPFV-DS.R.SIM man=DEF child-PL
ph-dé-·n
seeing-carry-IPFV
'While the woman cooked, the man watched the children.'

e. mé·n=da ʔ=doma mú·t̲u ča-lúš-či-**ba** qʰabé
so=at COP=HS 3.PAT sitting-slip-SML-SS.R.SEQ rock
ʔ=mi-l ʔ-t̤és-am
COP=there-on by.gravity-fall-down
'So then he$_i$ slipped and (he$_i$) fell onto the rock.'

f. bal há=w mčá-·č-'=**li** bal mú·t̲u
this mouth=in throw.PL-SML-PFV=DS.R.SEQ this 3.PAT
há=w čóq-'
mouth=in sting-PFV
'He threw them [the berries] into his mouth (and) it [the bee] stung him.'

(Mithun 1993: 120, 124, 127, 131)

In (298a) the suffix *-hi* indicates that the clause is in irrealis and that the subject in the next clause is the same (he$_i$ and he$_i$ shows that the subject is coreferential). In (298b)

the clause is in irrealis and the subject in the next clause is different, as indicated by *-hla*. The clauses in (298c–f) are all in realis. In (298c) the subject is the same across all clauses, and the events are all simultaneous, as indicated by the suffix *-(i)n* (effectively indicating something like *The man*$_i$ (*he*$_i$) *fought with his wife,* (*he*$_i$) *kicked her and* (*he*$_i$) *slammed her to the ground*). In (298d) the events are again simultaneous, but the subjects differ between the clauses, as indicated by *-da*. In (298e) the events are sequential (he first slipped, then fell), and the subject is the same in both clauses, as indicated by *-ba*. Finally, in (298f) the events are again sequential but the subjects differ between the clauses, being first he$_i$ then it$_j$ (or, more accurately, '3.SINGular$_i$' and '3.SINGular$_j$'), as indicated by *-li*.

11.4 Complex clauses in sign languages

Sign languages may, just as spoken languages, form complex clauses that are either coordinated, subordinated or cosubordinated. Coordination in sign languages tends to be asyndetic (where the units are juxtaposed without any overt linker) although some syndetic constructions (where an overt linker coordinates the units) can also be found. ST, for example, does have a sign for 'and' (the sign PLUS) but may also form asyndetic coordinate constructions, essentially in the form of a list. Here the non-dominant hand typically maintains the list as a kind of discourse guide while the dominant hand signs the various entities of the coordinated clause.

ST (Sign Language: Sweden)

(299) (tap index) DIRECTOR (tap middle finger) CONSULTANT (tap ring finger)
ONE-LIST----------TWO-LIST----------THREE-LIST
SECRETARY INCLUDE.tap ring finger ALSO WORK AS INTERPRETER
----------THREE
'One director, one consultant and one secretary who will also function as an interpreter.' (Ahlgren & Bergman 2006: 53)

In (299) we have a multiple coordination. The three-dimensional modality of sign languages allows for a coordination strategy whereby each coordinated unit is kept track of in the form of a list on the non-dominant hand. The list can thus easily go up to five units. There are no overt coordinating linkers in the construction, as the so-called list buoys (see below) serve to indicate how the clauses relate to each other. This is what Liddell calls **buoys** because they "help guide the discourse by serving as conceptual landmarks as the discourse continues" (2003: 223). In the above construction we have an example of **list buoys** (i.e. pointers to help maintain a list). Liddell also identifies theme buoys, fragment buoys and pointer buoys. Buoys seem to be widespread among sign languages. For more details, see Liddell (2003: 223ff).

Very common is to have asyndetic coordination, i.e. juxtaposition of the coordinated entities:

LIU (Sign Language: Jordan)

(300) a. FATHER Ø MOTHER DEAF $INDEX_1$ NEG// SPEAK
'My father (and) mother aren't Deaf, they speak.'
b. $INDEX_1$ EXIST FISH Ø NEG-EXIST $INDEX_3$
'I have a fish (and) she doesn't.' (Hendriks 2008: 79, 122)

In (300a) the units FATHER and MOTHER are simply juxtaposed to indicate coordination. Likewise, in (300b) the two clauses 'I have fish' and 'she does not have (fish)' are merely juxtaposed to indicate coordination; none of the constructions have any overt linking marker. In some sign languages juxtaposition is used for conjunctive coordination while adversative coordination is syndetic. NGT does not have any sign for 'and', but does have a sign BUT (van Gijn 2004). Auslan also has syndetic adversative coordination:

Auslan (Sign Language: Australia)

(301) K-I-M LIKE CAT BUT P-A-T PREFER DOG
'Kim likes cats but Pat prefers dogs.' (Johnston & Schembri 2009: 213)

In (301) the disjunction BUT is overtly marking the coordination of the two clauses 'Kim likes cats' and 'Pat prefers dogs'.

Subordinate complement clauses are often simply juxtaposed without any overt subordinating marker:

NGT (Sign Language: the Netherlands)

(302) $INDEX_3$ BELIEVE [EARTH FLAT]
'S/he believes [that the earth is flat].' (van Gijn 2004: 74)

In (302) the clause 'that the earth is flat' serves as the complement to the verb BELIEVE. There is no overt subordination marker. One type of overt indication of subordination, however, is scope of negation. Typically the negation has scope over the entire main clause, which means that it also includes subordinate clauses. Consider the following example:

ASL (Sign Language: USA)

(303) _____neg_____
$INDEX_1$ WANT [$INDEX_3$ GO-AWAY]
'I didn't want [him to leave].' (Sandler & Lillo-Martin 2006: 299)

In (303) the negation has scope over the entire main clause. If the sentence would have consisted of two coordinated main clauses, the negation would have had to be repeated. This is, in a sense, an overt indication that the second clause is embedded

in, i.e. subordinate to, the first (matrix) clause. A translation to the effect of something like 'I didn't want him and he left' is therefore not possible.

In many sign languages subordination is marked with non-manual signing and other prosodic features, such as pauses. In IPSL, for example, the subordinate clause is indicated by wide-open eyes and raised eyebrows, a head nod at the last sign of the subordinate clause and a pause after the last sign of the subordinate clause (Zeshan 2003c).

IPSL (Sign Language: India, Pakistan)

(304) MAIN' AMI:R PAIDA:I$ DEAF HAI MAIN' MADAD
I rich born deaf EXIST I help
subord---------- subord----------
'If someone is rich, and if he has a deaf child, he will help (other deaf people).'
(Zeshan 2003c: 204)[184]

In (304) we have two subordinate conditional adverbial clauses, 'if someone is rich…' and 'if someone has a deaf child…', both of which are marked as subordinate with non-manuals. Notice that the two conditional clauses are coordinated by juxtaposition. The difference between types of subordinate clauses is not made explicit in IPSL. Thus the second subordinate clause in (304) may also translate into a relative clause, as in 'if he has a child who is deaf'.

Some sign languages have relative pronouns that are distinct from other pronouns. In LIS, for example, a relative clause may be marked not only with the non-manual subordination marking of raised eyebrows, but also with relative pronouns:

LIS (Sign Language: Italy)

(305) BOY_i CALL $PROREL_i$ LEAVE DONE
subord----------
'The boy [who called] left.' (Cecchetto et al. 2006: 952f)

In (305) we have two separate markers for the subordinate clause; first of all, the general non-manual subordination marker that is held throughout the subordinate clause, and secondly, the relative pronoun, glossed as PROREL, which is coreferential with the head NP in the matrix. The relative pronoun "is signed with the wrist bent toward the floor, the hand closed with the index finger extended and moving from left to right" (Cecchetto et al. 2006: 953) while the personal pronoun is signed in the space of the entity referred to.[185]

---

184. It is in IPSL common to use the first person for general statements, as in 'if I am rich' meaning 'if someone is rich'.

185. See Cecchetto et al. (2006) for a discussion on why the LIS relative clause constructions should be interpreted as correlative clauses.

Many sign languages have constructions whereby sequences of motion verbs describe a single event:

AdaSL (Sign Language: Ghana)

(306) **ENTER**-head **MOVE GO-ROUND** DIRECTIONAL-INDEX-down
'(The ball) entered him through his head and so he went down (through the drainpipe).' (Nyst 2007: 186)

In (306) the verb sequence ENTER, MOVE and GO-ROUND describe the single event of the ball entering and going through the head. In other words, the verb sequence does not describe three separate events of the ball first entering, then moving, then going around. Furthermore, the verbs in the sequence share arguments. These kinds of sequences also tend to form a prosodic unit, typically manifested in a consistent eye gaze throughout the sequence as well as a lack of pauses between the verbs of the sequence. Due to these characteristics, these kinds of constructions have been interpreted as serial verb constructions, more specifically **serial verbs of motion** (Supalla 1990). As the units in these constructions are neither subordinated to each other nor function independently, this constitutes an instance of cosubordination.

11.5 Summary

All languages have the capacity to combine units and clauses to form more complex constructions. Main clauses are clauses that function independently as complete sentences. Main clauses may consist of a matrix clause and its subordinate clause(s). A matrix clause is a clause minus its subordinated clause; it may or may not function independently as a main clause. Subordinate clauses are embedded in their matrix clause and cannot function independently.

Coordination involves linking together two or more main clauses. Coordination may be conjunctive, disjunctive, adversative or causal. Coordinated constructions often involve ellipsis, or clause reduction, where repeated material is deleted to avoid redundancy.

In subordination one (or several) clause(s) is (are) subordinated to, or embedded in, another clause. Complement clauses are subordinate clauses that function as arguments to their matrix. Adverbial clauses function as adjuncts to their matrix. Relative clauses function as modifiers of a noun or NP.

Cosubordination involves complex clauses where units are not embedded in each other, which makes them similar to coordinated clauses, but where the units also do not function independently as full grammatical sentences, which makes them similar to subordinated clauses. A serial verb construction is a form of cosubordination whereby a single event is expressed by multiple verbs that share grammatical features and arguments and form one prosodic unit. Clause chaining is a construction where multiple medial verbs form a chain that is embedded in their matrix clause. Chained clauses typically express switch reference.

Coordinated clauses are very often asyndetic in sign languages, although adversative clauses are more likely to be syndetic than conjunct clauses. It is very common for sign languages to make use of buoys, for example list buoys, which keep track of the coordinated units. Subordinated complement clauses are also typically juxtaposed, but many sign languages make use of non-manual subordination markers for adverbial and relative clauses. Cosubordination in the form of serial verbs, specifically serial verbs of motion, is widespread among sign languages.

11.6 Keywords

adverbial clause
clause chaining
complement clause
coordination
cosubordination
ellipsis
embedding
juxtaposition
main clause
matrix clause
medial clause
relative clause
serial verbs
subordination
switch reference
syndetic coordination

11.7 Exercises

1. How do pidgin and creole languages pattern with respect to the position of relative clauses compared to non-creole languages?
2. Which of the following sentences would you consider an example of a serial verb and which an example of a medial clause? Why?

   Mauwake (Trans-New Guinea (Madang): Papua New Guinea)

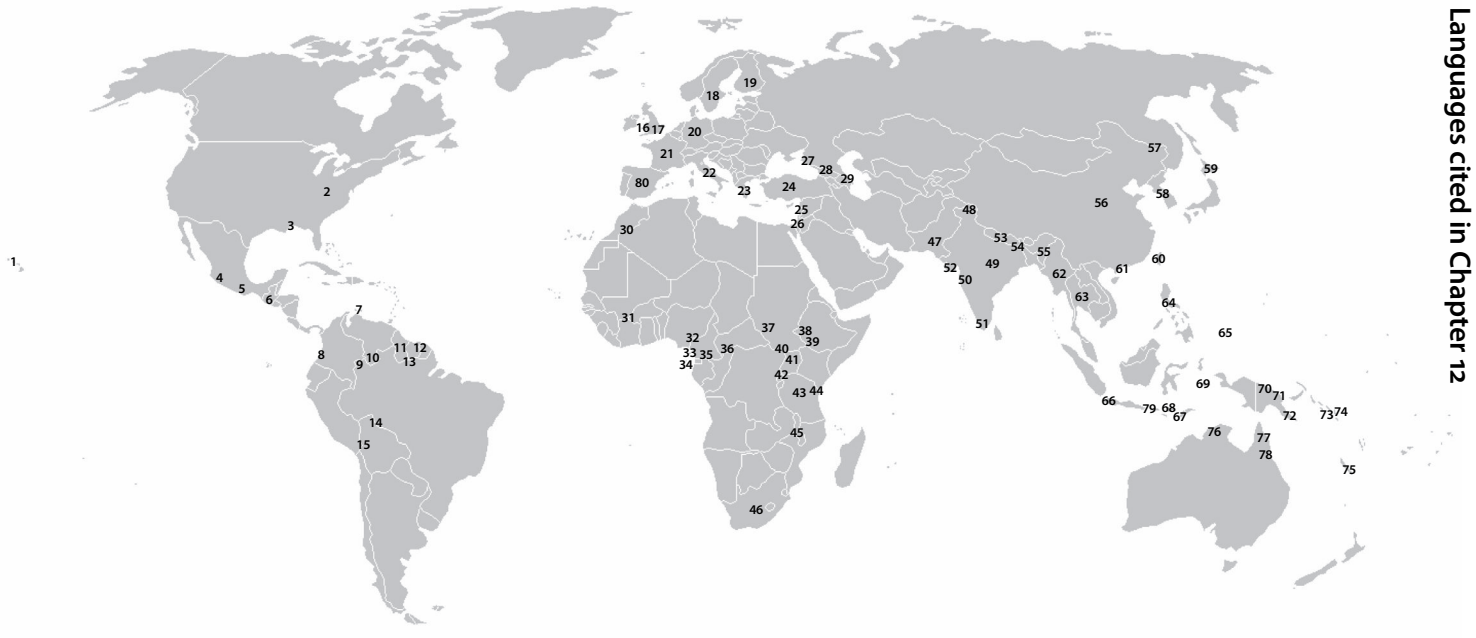
   a. wiawi ikiw-ep maak-eya ne wiawi=ke maak-e-k…
      3.father go-SS.SEQ tell-2/3SG.DS ADD 3.father=CF tell-PAST-3SG
      'She went to her father and told him, and her father told her…'

   b. mika-fien-ikiw-o-k
      hit-push.aside-go-PAST-3SG
      'He went on countering (an attack)' (Berghäll 2010: 173, 300)

3. What is **switch reference**?
4. What are sign language **buoys**?
5. Is the following statement true or false? Motivate your answer.

   If a language has VO word order then it is also likely to have NRel word order.

Languages cited in Chapter 12



1 Hawai'i Creole English
2 American Sign Language
3 Biloxi
4 Purépecha
Tzutujil
5 Mixtec, Chalcatongo
7 Papiamentu
8 Epena Pedee
9 Tariana
Tucano
10 Warekena
11 Macushi
12 Ndyuka
13 Tiriyo
14 Wari'
15 Puquina
16 Welsh
17 British Sign Language
English
18 Swedish
Swedish Sign Language
19 Finnish
Finnish Sign Language
20 German
21 French
22 Italian
Italian Sign Language
Latin
23 Greek Sign Language
24 Turkish
Turkish Sign Language
25 Lebanese Sign Language
26 Israeli Sign Language
27 Abkhaz
28 Kabardian
29 Lezgian
30 Tashlhiyt
31 Supyire
32 Lokaa
33 Pichi
34 Angolar
Principense
35 Makaa
36 Sango
37 Dinka, Agar
38 Sheko
39 Zayse
40 Juba Arabic
41 Lango
42 Nkore-Kiga
43 Burunge
44 Ma'a/Mbugu
45 Chichewa
46 Afrikaans
47 Indo-Pakistani Sign Language, Karachi
48 Kashmiri
49 Indo-Pakistani Sign Language, Delhi
50 Korlai
51 Malayalam
52 Diu Indo-Portuguese
53 Nepali
54 Dumi
Maithili
55 Naga Pidgin
56 Chinese Sign Language
Mandarin
57 Chinese Pidgin Russian
58 Korean
South Korean Sign Language
59 Ainu
60 Taiwanese Sign Language
61 Hong Kong Sign Language
62 Burmese
63 Thai Sign Language
64 Chabacano: Caviteño
Chabacano: Zamboanga
65 Palauan
66 Batavia Creole
Indonesian
67 Sawu
68 Manggarrayi
69 Ambon Malay
70 Yimas-Arafundi Pidgin
71 Tauya
72 Orokaiva
73 Lavukaleve
74 Toqabaqita
75 Tinrin
76 Gaagudju
Waray
77 Guugu Yimidhirr
78 Dyirbal
79 Kata Kolok
80 Spanish Sign Language

Chapter 12

Speech acts and politeness

Languages are universally used in interaction between individuals. When we engage in any kind of communication we do so with the purpose of giving information, requesting information, or getting people to do something. This chapter first deals with speech acts (12.1) and the main sentence types found in all known languages, both spoken and signed: declaratives (12.1.1), interrogatives (12.1.2) and imperatives (12.1.3). I then briefly discuss issues relating to politeness (12.2) and give a few examples of some devices languages make use of to uphold the necessary levels of politeness, such as through indirect speech acts (12.2.1) and honorifics (12.2.2). Section 12.3 gives an overview of topics related to speech acts and politeness in sign languages.

12.1 Speech acts

We should never forget that language is used as a medium of communication between individuals. While a language may be used for innumerable purposes, all known languages are capable of serving at least three functions: to pass on information, to glean information, and to issue commands. In fact we rarely just make an utterance for no purpose whatsoever.[186] As soon as we use language for any of these three core purposes of asserting something, asking something or commanding something, we engage in a **speech act**, that is, we perform an act by making an utterance.

The three basic sentence types, or types of speech act, that seem to be universal to human language are declaratives, interrogatives and imperatives. Often we may identify further sentence types, such as prohibitives and optatives, as subcategories of these basic speech act types (König & Siemund 2007). A **declarative** is basically an assertion, a statement made in order to pass on information. An example of a declarative sentence is *You are sitting on the floor*. An **interrogative** is a request for

---

186. This actually also holds true for so called **phatic communion**, where utterances are made not to convey information, but for social reasons, such as maintaining social contact or in order to avoid long silences – in other words, when we engage in small talk. Here too the act serves a purpose, namely the purpose of comfortable social interaction. See, for example, Laver (1975). But see also Urbanová (2007) for a discussion on why phatic communion and small talk are two separate (even if interrelated) concepts.

information, posed as a question with the intention of gleaning information. Examples of interrogatives are *Why are you sitting on the floor?* and *Are you sitting on the floor?* (the difference between these two kinds of questions will be discussed below). An **imperative** is a command, uttered with the intention of getting someone to do something. An example of an imperative is *Sit on the floor!*

The following will discuss how the three sentence types mentioned above are coded grammatically among languages. It is beyond the scope of this chapter to give a comprehensive discussion on all the subtleties involved in the myriad ways speakers may use their language for various purposes. For an accessible overview on theories of speech acts, see Sadock (2006). Geis (2006) provides a thorough discussion on how speech acts are employed in natural conversation. For a very accessible discussion on the typology of speech act distinctions, including minor sentence types such as exclamatives and answers to questions, see König & Siemund (2007). For a typological overview of non-declarative speech acts, see Givón (2001b: 287ff).

12.1.1 Declaratives

Declaratives are usually the least marked sentence type and are typically used for statements, such as *The puppy is playing with a ball*. Declarative sentences can be either affirmative or negative, where the former expresses an assertion about something and the latter negates (reverses) an assertion. For instance, an assertion like *The man is walking* (or *The puppy is playing with a ball* or any other plain statement) may be negated to *The man is not walking* (or *The puppy is not playing with a ball*), thus reversing the assertion of the affirmative.

12.1.1.1 *Affirmative declaratives*

Affirmative declaratives are typically used for descriptive speech acts, such as asserting something, describing something, complaining or bragging about something, predicting or promising something, and so on. It is typically the most frequent sentence type, it is typically the least restricted in its distribution, and it may typically combine with all the TMA options of the language.

While affirmative declaratives are most commonly unmarked as a sentence type, this is by no means an absolute universal. Sheko is an example of a language where affirmative declarative sentences must also be marked:

Sheko (Afro-Asiatic (North Omotic): Ethiopia)

(307) a. koṃ-s maak-ab-ara íʃì-ʃe-ke
chief:DEF-M tell-REL-ACC 3PL-forget-DECL1
'They forgot what the chief told them.'

b. k'áy-ē, gob sats'-á-ke
rise-IMP sky become.light-3MS-DECL1
'Stand up, it has become light/the sun came up.' (Hellenthal 2007: 22)

In (307a) we have a straightforward statement. The declarative marker *-ke* obligatorily attaches to the verb; something like **koṃs maakabəra íʃìʃe* (without the declarative marker) would not be acceptable. In (307b) we have an example of an utterance containing an imperative (*k'áyē* 'Stand up!') and a declarative (*gob sats' áke* 'The sun has risen.'). Both sentence types receive their own specific markers, *-ē* and *-ke* respectively. It should be noted that in addition to the declarative *-ke*, Sheko has a second declarative marker *-me* which appears only in irrealis assertions, both of which are in paradigmatic opposition to other modal markers such as imperative and optative (Hellenthal 2007). It could therefore conceivably be argued that *-ke* should be analysed as a realis marker and *-me* as an irrealis marker rather than declarative markers. In Tauya, however, we have a declarative which appears in both realis and irrealis propositions and which may combine with modal markers:

Tauya (Trans-New-Guinea (Madang): Papua New Guinea)

(308) a. mene-a-te mepi ʔatou-a-ʔa
stay-3SG-DS come.down arrive-3SG-DECL[187]
'She stayed and he came down and arrived.'

b. ya-ni tei-mene-amu-ʔa
1SG-ERG catch-STAT-1SG.FUT-DECL
'I will have it.'

c. ʔei-ra mene-a-rafo-ʔa
there-TOP stay-3SG-DUB-DECL
'Maybe he's there.' (MacDonald 1990: 208–9)

In (308a) we have a plain statement; the declarative marker *-ʔa* necessarily attaches to the finite verb. In (308b) *-ʔa* is still necessary, even though we have a prediction, that is, a statement that pertains to the future, which is an irrealis proposition. Thus *-ʔa* is not a realis marker. In (308c) *-ʔa* combines with the dubitative modal marker (*-rafo-*), showing that it is not a modal marker in paradigmatic opposition to other modal markers.

12.1.1.2 *Negative declaratives*

A **negation** reverses the truth of a proposition. For instance, the declarative *Mary is drinking tea* states the actuality that Mary is drinking tea. Adding a negative marker reverses this actuality: *Mary is not drinking tea* states that whatever it is that Mary may be doing, she is not drinking tea. All known languages have a way of forming a negative declarative, but the strategies may differ. Languages may also have different types of negation. A **clausal negation** is a negation that negates an entire clause, giving a negative counterpart to an affirmative declarative, as in *Mary is not drinking tea* (the

187. *-ʔa* is glossed as 'indicative/unmarked' in the source.

negative declarative counterpart of the affirmative *Mary is drinking tea*). A **constituent negation** negates only a particular constituent in the clause, such as NP negation as in *No tea could be found*, where the negative *no* only refers to the NP *tea* and not to the entire clause. Languages may also have **negative pronouns**, as in *Nothing could be found*, and **negative adverbs**, as in *Mary never drinks tea*. This section will deal specifically with clausal negation in declarative sentences. For a very accessible overview of negation, see Miestamo (2007). See also Givón (2001a: 369ff).

The negation can have different **scope** over the sentence. Scope essentially refers to that section which is affected by the meaning of a particular form. In English, for example, the negation typically has scope over everything that comes after it. Compare *I deliberately didn't step on the snail* with *I didn't deliberately step on the snail*. In the former sentence it was a conscious and premeditated act to not step on the snail. In the latter sentence I did step on the snail, but the act was accidental.

In all known languages clausal negation is realized through morphology one way or another: "[t]here are no known instances of languages in which negation is realized by a change in word order or by intonation, and all languages have negative morphemes" (Dryer 2011g). Languages tend to have either negative particles or negative affixes – of the 1159 languages in Dryer's (2011g) database on negative morphemes, 502 (or 43.3%) form negative declarative sentences through negative particles, such as English with the particle *not* or Gaagudju with the particle *gaayu*.[188] This strategy is spread all over the world.

Gaagudju (Australian (Gaagudju): Australia)

(309) **gaayu** i-n-yii-ngi
NEG 3I-IRR-GO-PAST
'He didn't go.' (Harvey 2002: 322)

The second largest group in Dryer's database, with 396 languages (34.2%), forms negative declarative sentences by way of a negative affix. This is also a strategy spread all over the world, except in the largely isolating Southeast Asian languages. Chichewa is an example of such a language, with its negative prefix *si-* (or *s-* if it precedes a vowel).

Chichewa (Niger-Congo (Bantoid): Malawi, Mozambique, Zambia, Zimbabwe)

(310) Mkângo **s**-ú-ku-wá-phwány-a maûngu.
III.lion NEG-III.SM-PRES.-VI.OBJ-smash-FV VI.pumpkins
'The lion is not smashing them, the pumpkins.' (Mchombo 1998: 503)

188. Gaagudju has another negative particle, *ibárdbi*, which "does not deny the occurrence of an event or the existence of an entity, but instead means something like 'it is not the case that X can be predicated of Y'" (Harvey 2002: 321).

The remaining types of strategies all make up proportionally rather small groups of languages. In 47 languages (4.1%) in Dryer's sample the negative declarative is formed through a type of auxiliary verb which inflects for person and number in the same way as other auxiliaries. Finnish is an example of such a language, as shown in (85) repeated here for convenience:

Finnish (Uralic (Finnic): Finland)

(311) **en** tule kotiin
NEG.1SG come home.ILL
'I won't come home.' (Sulkala & Karjalainen 1992: 115)

In (311) the auxiliary *en* is inflected for person and number, in this case first person singular. While this is a comparatively rare strategy, languages with auxiliary verb negatives are found in most parts of the world, but are especially prominent in northern Eurasia in a belt stretching from Finland to the Russian Far East, along the North American Pacific coast and in a cluster in Central Africa. Auxiliary verb negatives are completely absent in Australia, Europe (except for Finland) and the Middle East in Dryer's sample.

In 73 languages (6.3%), clustered mainly in West Africa, Southeast Asia and the South Pacific islands, it is not possible to determine if the negator is an auxiliary or a particle. For example, if a language has little or no morphology and the verb carries no inflection, a negative word may be an auxiliary without any inflection (just like the other verbs), but it may also be a particle. In Tinrin, for example, the negative is not inflected, but neither is the verb:

Tinrin (Austronesian (Oceanic): New Caledonia)

(312) nrâ **see** tròa warra nrâ toni
3SG NEG arrive yet SM PN
'Tony has not arrived yet.' (Osumi 1995: 224)

In cases like Example (312) it is not possible to determine whether the negative word *see* should be considered a particle or an auxiliary verb, although for some languages there might be syntactic criteria that indicate how the negative is best classified.

A small group of 21 languages (1.8%), found in all parts of the world in Dryer's sample except Europe, the northern Eurasian mainland and Southeast Asia, have both a negative affix and a negative word. In Orokaiva, for example, negation is expressed with the suffix *-ae* except for those clauses that are in the future tense or habitual aspect, for which the particle *mane* is used:

Orokaiva (Trans-New Guinea (Binanderean): Papua New Guinea)

(313) a. na pamb-**ae**
1SG go-NEG
'I didn't go.'

(313) b. na **mane** pamba-so-n-a
1SG NEG go-FUT-1SG-IND
'I will not go.' (Larsen 1977: 21)

Finally, 120 languages (10.4%) in Dryer's sample have double negation, i.e. two simultaneous negative morphemes.[189] In standard French, for example, the negative declarative sentence is marked by *ne … pas*:

French (Indo-European (Romance): France)

(314) je **ne** veux **pas** manger
1SG.NOM NEG want.1SG.PRES NEG eat
'I don't want to eat.' (source: personal knowledge)

Because colloquial French regularly drops *ne* (giving sentences like *Je veux pas manger* 'I don't want to eat'), it is coded as a language with negative particles rather than with double negation in Dryer (2011g). In Makaa this double negation is expressed through the parafix *a-…(H)-ɛ* and a high tone (glossed as (NEG.H) here) on the verb stem.

Makaa (Niger-Congo (Bantoid): Cameroon)

(315) mə̀ **à**-cál-**ɛ́** mə́-lə́ndú
1SG NEG-cut(NEG.H)-NEG VI-palm.tree
'I do not cut down palm trees.' (Heath 2003: 346)

This strategy is especially common in sub-Saharan Africa and Papua New Guinea, but can also be found in both North and South American languages.

Languages may display secondary strategies of negation that need to be employed alongside the primary strategy. For instance, in Makaa (Example (315) above) we find not only the double negator affixes, but also an obligatory change in tone: the verb stem gets a high tone together with the parafixes. Another secondary strategy in negative declaratives may be change of word order, as in Lokaa, where affirmative declaratives are typically SVO while negative declaratives are typically SOV:

Lokaa (Niger-Congo (Cross River): Nigeria)

(316) a. Úbì ó-kpèèyì kò-póó
PN I.AGR-sell XI-cup
S V O
'Ubi sold a cup.'

189. The term 'double negation' is sometimes used for such constructions as *I've **not** said **nothing*** (where two negative forms are used in the same clause).

b. Úbì kò-póó òó-kpèèyì
PN XI-cup NEG/I.AGR-sell
S O V
'Ubi didn't sell a cup.' (Baker 2005: 126)

Notice that the tone of the verb stem *-kpèèyì* 'sell' remains the same in both sentences. In other words, the negative only affects the word order, and not the tone. Yet another secondary effect of negation may be the number of TMA distinctions allowed; in languages where TMA distinctions are affected by negation, there are typically fewer distinctions in the negative than in the affirmative. In Burmese, for example, the punctual aspect and the realis/irrealis modal distinctions are lost in the negative:

Burmese (Sino-Tibetan (Burmese-Lolo): Myanmar)

(317) a. thu yau' pi
3SG reach PUNCT
'(S)he has arrived.'
b. thu **ma**-yau' **hpu:**
3SG NEG-arrive NEG
'(S)he hasn't arrived.'
c. tin hpei a ye' thau' tha-la
PN liquor drink R-Q
'Does Tin Pe drink liquor?'
d. tin hpei a ye' **ma**-thau' **hpu:**
PN liquor NEG-drink NEG
'Tin Pe doesn't drink liquor.'
e. thu pyo: ma-la
3SG speak IRR-Q
'Will (s)he speak?'
f. thu **ma**-pyo: **hpu:**
3SG NEG-speak NEG
'(S)he won't speak' (Soe 1999: 145–7)[190]

The punctual aspect in (317a) as well as the realis and irrealis mood in (317c and e) respectively are all neutralized with the negative, as can be seen in (317b, d and f).

190. The transcription and glossing in Soe (1999) differs somewhat from that of others. In Cornyn (1944), for example, the negative *ma- … hpu:* is transcribed *ma- … phû/bû*; the punctual marker *-pi* is transcribed *-pí/-bí* and analysed as denoting that "the action or condition has already begun" (12); the realis marker *tha* is transcribed *-té/-dé* and glossed as "without reference to time, which may be present or past" (12); the irrealis marker *ma* is transcribed *-mé* and glossed as a future tense marker (12). Notice also that Soe does not mark tone.

These secondary strategies are what may be called asymmetric negations. In symmetric negations "the structure of the negative is identical to the structure of the affirmative, except for the presence of the negative marker(s)" (Miestamo 2011b) while in asymmetric negations the structure differs from the affirmative one way or another. Of the 297 languages in Miestamo's (2011b) database, 114 (or 38.4%) have a symmetric negation strategy while 53 (17.8%) have asymmetric negation. The largest group, with 130 languages (43.8%), has both symmetric and asymmetric negation, in that it is symmetric in some contexts but not in others. English is an example of a language classed as having both symmetric and asymmetric negation due to the asymmetry displayed with emphatic sentences (see Miestamo 2011a). Sentences like *He has eaten* versus *He has not eaten* exhibit symmetric negation. However, asymmetry is exhibited with respect to the distinction between emphatic and non-emphatic sentences (such as *He sings* versus *He does sing*, where the latter is pragmatically marked for emphasis), which is lost in negation (both have to be negated as *He does not sing*).

12.1.2 Interrogatives

All languages have some way of gleaning information, that is, of forming **interrogatives** with the purpose of extracting information. Depending on what kind of information is requested, interrogatives may take the shape of different kinds of questions. If the information sought after is merely confirmation or disconfirmation of the truth value of an utterance (essentially through the answers *Yes* or *No*), we form polar questions. If we need more elaborate information, which cannot be provided by a mere *Yes* or *No*, we form content questions. An example of a polar question is *Will you come home today*? where the answers may be either *Yes* or *No* (although various intermediate answers may also be permitted, such as *Maybe* or *I don't know*). An example of a content question is *When will you come home*? where more elaborate information would be needed than *Yes* or *No* to constitute an acceptable answer.[191]

12.1.2.1 *Polar questions*

With **polar questions** (also called **yes-no questions**) the expected answer is typically *Yes* or *No*, although, as mentioned, there are usually a few other alternatives available (such as *Perhaps*). Most languages have specific strategies for forming polar questions, often a combination of strategies.

It is very common for languages to have a distinct intonation pattern for polar questions. Often a polar question has a rising intonation, as is the case in English (try

191. But see Huddleston (1994), who makes a clear distinction between interrogatives and questions in that interrogatives are a syntactic type of sentence, contrasting with declaratives, imperatives and exclamatives, while questions are a semantic category defining the set of answers required.

listening to yourself first saying *You are home* then *Are you home?*), but this is not a universal. Hawai'i Creole English, for example, has a falling intonation for polar questions (source: own fieldwork). In a number of languages intonation may be the only device to mark a polar question. Of 954 languages in Dryer's (2011s) database on polar questions 173 (18.1%) mark polar questions through intonation only. These languages are spread over the world, but are especially common in sub-Saharan Africa, Australia and Papua New Guinea, as well as South and Meso-America. Italian is an example of such a language.

Italian (Indo-European (Romance): Italy)

(318) a. Laura viene con ↘noi (statement)
PN come.3SG.PRES.IND with 1PL
'Laura is coming with us.'

b. Laura viene con ↗noi (question)
PN come.3SG.PRES.IND with 1PL
'Is Laura coming with us?' (source: personal knowledge)

In (318a and b) the only difference between the two sentences is that the declarative statement has a falling intonation (indicated by ↘) while the interrogative polar question has a rising intonation (indicated by ↗). It should be noted that intonation as a strategy is underrepresented in Dryer's map, since a language has only been coded for that strategy if there are no other grammatical means for forming a polar question in the language. However, "[i]n some languages, intonation may be the most common means of indicating a polar question, but if some other method is used a minority of the time, then the language is shown on the map according to that method" (Dryer 2011s). It is common for languages to combine intonation with any of the strategies discussed below.

A well known, but actually quite rare strategy for marking polar question is through change of word order. This is found in only 13 languages (1.4%) in Dryer's database, almost all of them clustered in Western Europe, except for two in insular Southeast Asia (Manggarayi (Austronesian (Central Malayo-Polynesian): Indonesia) and Palauan) and one in South America, Warekena (Arawakan (Arawakan): Venezuela). Swedish is an example of a language that employs inverted word order to form polar questions:

Swedish (Indo-European (Germanic): Sweden)

(319) a. han kommer
3SG.M come.PRES
'He is coming.'

b. kommer han
come.PRES 3SG.M
'Is he coming?' (source: own knowledge)

English is also an example of such a language (compare the translations of Example (319) above, where the inflected verb moves to sentence initial position in the polar question).

A very rare strategy indeed is to mark polar questions through the absence of a declarative marker. This is found in only four languages (0.4%) in Dryer's database, Dinka, Kabardian (Northwest Caucasian (Northwest Caucasian): Russia), Puquina and Zayse (Afro-Asiatic (North Omotic): Ethiopia).

Puquina (Isolate: Bolivia)

(320) a. cuhaña-qu-ench
believe-1S.SG-DECL
'I believe.'
b. cuhaña-pi
believe-2S
'Do you believe?' (Adelaar & Muysken 2004: 354)

In (320a) the obligatory declarative marker *-(e)nch*/*-(a)nch* suffixes to the verb, while in (320b) the absence of this marker indicates that the utterance is a polar question.

A rather more common strategy is to have a special verbal inflection signalling the polar question. This is found in 164 languages (17.2%) in Dryer's database, also spread over the world, but especially common in mainland South and Southeast Asia as well as South and North America. Korean is an example of such a language, with its question marker suffix *-ni* or *-nyi*, as is Burmese (see Example (317c and e) above).

Korean (Isolate: N, S Korea)

(321) pi ka o-**ni**/o-(nu)-**nyi**
rain NOM come-Q/come-IND-Q
'Is it raining?' (Sohn 2001: 269)

The by far most common strategy in Dryer's database is to have **question particles**, which may either be a free particle or a clitic added to the declarative sentence. This is found in 584 languages (61.2%) spread all over the world except in Western Europe. Ainu is an example of a language with a question particle, as is Supyire (see Example (127) above).

Ainu (Isolate: Japan)

(322) pirka-p ne **ya**
rich-person be Q
'Is (he) a rich person?' (Shibatani 1990: 24)

In (322) the particle *ya* signals that the utterance is a question. In Ainu this particle is sentence final. This is the most common pattern for languages with question particles (Dryer 2011v). Question particles may also be sentence initial, as in Tzutujil:

WHISPERED STATEMENTS

In Nkore-Kiga, the only difference between the statement and the question is how the final syllable is pronounced. In declaratives the final syllable is whispered (indicated through superscript here), while in interrogatives it is voiced. This strategy can also be found in Burunge (Afro-Asiatic (Southern Cushitic): Tanzania) (Kiessling 1994).

Nkore-Kiga (Niger-Congo (Bantoid): Uganda)

a. n'-omushaij[a]
   AC-man
   'It is a man.'

b. n'-omushaija
   AC-man
   'Is it a man?' (Taylor 1985: 6)

Tzutujil (Mayan (Mayan): Guatemala)

(323) **la** n-at-war-i
Q INCOMPL-2SG-sleep-IVNPFS
'Are you going to sleep?' (Dayley 1981: 334)

Question particles or clitics may also appear in other positions in the sentence. For more details, see Dryer (2011v). Languages may also have **tag questions** (or **interrogative tags**), which can be seen as a type of question particle but which "contribute a certain bias by raising expectations toward either a positive or a negative answer" (König & Siemund 2007: 296). English has tags where the negative tag presupposes a positive answer and a positive tag a negative answer:

(324) a. You do know it, don't you? (Expected answer: Yes, I do.)
b. You don't know it, do you? (Expected answer: No, I don't.)

Tags tend to occur at the end of the sentence irrespective of the basic word order of the language. They are sometimes considered a shortened version of a different strategy, the **disjunctive-negative structure**, also known as '**A-not-A-construction**', which is especially common in Chinese languages, such as Mandarin, but can also be found elsewhere.

Mandarin (Sino-Tibetan (Chinese): China)

(325) tā chī-bu-chī píngguǒ
3SG eat-not-eat apple
'Does s/he eat apples?' (Li & Thompson 1990: 828)

In (325) the polar question is formed by giving the verb (*chī* 'eat') twice, once positive and once negated. It may be that tags originate from A-not-A-constructions, and that question particles, in turn, originate from tags and the A-not-A-construction (see Harris & Campbell 1995: 294ff for a discussion on the origin of question particles).

Notice, however, that some languages use the A-not-A-construction together with a question particle:

Turkish (Altaic (Turkic): Turkey)

(326) ahmet sinema-ya git-ti mi git-me-di mi
PN cinema-DAT go-PAST Q go-NEG-PAST Q
'Did Ahmet go to the cinema (or didn't he)?' (Kornfilt 2003: 8)

In Turkish the A-not-A-construction is optional and, when used, combines with the question particle (which is not optional).

A few languages in Dryer's (2011s) database, 15 (1.6%) to be exact, have both question particles and an interrogative verb morphology. Kashmiri, for example, has a question particle *k'a:* and a verbal interrogative suffix *-a:*, which may either occur alone or together.

Kashmiri (Indo-European (Indic): Pakistan, India)

(327) **(k'a:)** tsɨ gatsh-kh-**a:** paga:h garɨ
Q 2SG go-2SGPS-Q tomorrow home
'Will you go home tomorrow?' (Wali & Koul 1997: 5)

One single language in Dryer's database, Chalcatongo Mixtec, lacks any kind of differentiation between statements and polar questions.

Chalcatongo Mixtec (Oto-Manguean (Mixtecan): Mexico)

(328) ñábaʔa=ró librú=ro
have=2 book=2
'You have your book.' / 'Do you have your book?' (Macaulay 1996: 126)

The utterance in (328) can be interpreted either as a statement or a question; there is no overt marking and no intonational difference between the meanings. It is safe to assume that this is extremely rare cross-linguistically.

12.1.2.2 *Content questions*

Content questions (also called **question-word questions**, **information questions**, ***wh*-questions**[192] and **constituent interrogatives**) contain an interrogative phrase and demand a specific answer containing other information than just a confirmation or nonconfirmation. In other words, a simple *Yes* or *No* is not an adequate answer to a content question. An example of a content question is *What are you reading?*, where the answer would have to give specific information, such as, for example, *Shakespeare*

192. The term ***wh*-questions** derives from the fact that many of the English questions words start with *wh-* in writing: ***wh**o(m)*, ***wh**at*, ***wh**ere*, ***wh**y*, ***wh**en*, ***wh**ich* (but cf. ***h**ow*, which is also a question word). However, this only reflects English and while it may be useful as a mnemonic device for speakers and students of English, it is not very useful for cross-linguistic purposes.

or *The book about moths* or something similar. An **interrogative phrase** may consist of only a question word (e.g. *What are you reading?*) but may also consist of more than that, as in *Which book are you reading now?* or *Which of the new books will you be taking along?* where the question word plus the noun phrase (*which book* and *which of the new books* respectively) constitute the interrogative phrase.

The interrogative phrase replaces the constituent asked about, thereby indicating what kind of information is being requested. Thus the subject in a declarative like *John*$_S$ *read the book* is, in a corresponding content question, replaced by the interrogative phrase *who*: *Who*$_S$ *read the book?* Likewise, the object may be inquired about, as in *What*$_O$ *did John read*? Notice that the interrogative phrase in English is by default in sentence initial position, leaving a "gap" where the relevant constituent would have been in a declarative sentence (compare *John read* [*the detective story*]$_O$ with [*What*]$_O$ *did John read* ___ *?*). This movement of the constituent to sentence initial position is sometimes called **fronting**. To require the interrogative phrase to be in sentence initial position, as English does, is rather common cross-linguistically: of 901 languages in Dryer's (2011u) database on the position of the interrogative phrase, 264 (29.3%) obligatorily place it sentence initially. Much more common, however, is to allow the interrogative phrase to occur elsewhere in the sentence. This is true for 614 languages (68.1%) in Dryer's database, such as Dumi, where the interrogative phrase is in the same position as the constituent asked about would have been in a declarative sentence.

Dumi (Sino-Tibetan (Bodic): Nepal)

(329) a. an-a **mwo:** a-dzɨ-t-a
2SG-ERG what MS-eat-NPST-2SS
S O V
'What are you eating?'

b. khi:bɨ-ʔa aŋ a-ka:ts-ə
dog-ERG 1SG MS-bite-1SG
S O V
'The dog bit me.' (van Driem 1993: 62, 86)

The question word *mwo:* 'what' in (329a) is an object and placed in the same position as an object in a declarative sentence (b). In most of the languages that allow the interrogative phrase in other positions than sentence initial the interrogative phrase occupies the same position as the equivalent constituent would in a declarative sentence; it remains ***in situ***.[193] The remaining 23 (2.6%) languages in Dryer's database exhibit a mixed behaviour, in that some interrogative phrases occur sentence initially and some do not.

193. Notice, however, that sentence initial position of an interrogative phrase may also be a case of *in situ* if that is where the constituent would be in declarative sentences, as is the case in English if the interrogative phrase is a subject.

All known languages have a set of words that are used for content questions, but not all languages have specific question words. In Wari', for example, a content question is formed by placing a form of the demonstrative *ma'* in initial position and leaving a corresponding gap in the clausal syntax.

Wari' (Chapacura-Wanhan (Chapacura-Wanhan): Brazil)

(330) a. ma' co pa' nana hwam pain
that.PROX.HEARER INFL.M/F.RP kill 3PL.RP fish PREP.3N
S V O
xec ca' ne
day this.N REC.PAST
'Who killed fish this morning?'

b. mam wac 'ina-on hwam pain quit
INST cut 1SG.RP-3SM fish PREP.3N knife
V S O
'I cut the fish with the knife.' (Everett & Kern 1997: 18, 25)

The basic word order of Wari' is VSO (as shown in Example (330b)). Content questions are formed by fronting the demonstrative *ma'* 'that (proximal to hearer)' with a corresponding inflection morpheme (in this case *co*). However, since content questions in Wari' are by default formed by fronting the demonstrative, it may still be argued that it has a specific set of words characteristic for content questions.

Typical question words are interrogative pronouns (such as *who* or *what*), interrogative adverbs (such as *where* or *how*), interrogative determiners (such as *which*), interrogative quantifiers (such as French *combien* 'how.many'), and so on. Languages may also have interrogative verbs, that is, verbs that in themselves function as a question word. Lavukaleve is an example of such a language:

Lavukaleve (Solomons East Papuan (Lavukaleve): Solomons Islands)

(331) me-kalem **vasia**-m
2PL-father(M) be.where-SG.M
'Where is your(pl) father?' (Terrill 1999: 413)

The verb *vasia* 'be.where' in (331) is in itself a question word, not properly translatable into English, since English does not have interrogative verbs. The semantics of interrogative verbs is often something like 'be.who/what' or 'do.what'/'what.happened' but may also be something like 'be/do.how', 'say.what' and 'be/go.where'. For more on the typology of interrogative verbs, see Hagège (2008). An unusual question word meaning 'which part of X's body' can be found in Toqabaqita:

Toqabaqita (Austronesian (Oceanic): Solomon Islands)

(332) **tafa**-mu n-e fii
which.part.of.body-2SG.PERS FOC-3SG.NFUT hurt
'Which part of your body hurts?' or 'Where (in your body) does it hurt?'
(Lichtenberk 2007: 604)

The strategies for deriving polar questions discussed above are often also used with content questions. However, with content questions these strategies tend to be optional rather than obligatory. Furthermore, content questions generally do not allow tags or A-not-A-constructions.

12.1.3 Imperatives

All known languages have some kind of strategy for expressing directive speech acts, or **imperatives**, i.e. speech acts meant to get the addressee to perform some kind of action. Imperatives are typically used to issue commands (*Sit down!*), requests (*Please switch off the light.*), entreaties (*Let me try it out!*), advice (*Don't bike without a helmet!*), warnings (*Watch out for the dog!*), instructions (*Stir a spoonful of port in the sauce.*), wishes (*Have fun!*), and so on. Imperatives may be positive or negative, in that the former directs the addressee towards an action that should be done, while the latter prohibits the addressee from an action. For a very detailed study on the forms and functions of imperatives and commands, see Aikhenvald (2010), which this section relies heavily on.

12.1.3.1 *Positive imperatives*

Positive imperatives, usually simply termed imperatives, are used to initiate action. In addition to the different morphological strategies for marking imperatives discussed in 8.5.3.1, it should be noted that some languages, such as Abkhaz, make a gender distinction in singular imperatives.

Abkhaz (Northwest Caucasian (Northwest Caucasian): Georgia)

(333) a. **w**-c'a (male addressee)
M-go
'Go!'
b. **b**-c'a (female addressee)
F-go
'Go!'
c. **šw**-c'a (plural addressees)
PL-go
'Go!' (Wolfgang Schulze, p.c.)

In (333) the imperative takes a gender dependent prefix in the singular (*w-* 'masculine' and *b-* 'feminine' respectively). This distinction is, however, neutralized in the plural, as shown in (333c). In Tashlhiyt the gender distinction appears in the plural but not in the singular. This is very rare cross-linguistically.

Tashlhiyt (Afro-Asiatic (Berber): Morocco)

(334) a. du: (said to one male or female addressee)
'Go!'

b. du:-**iat** (said to more than one male addressee)
go-M.PL
'Go!'

c. du:-**iamt** (said to more than one female addressee)
go-F.PL
'Go!' (Applegate 1955: 42)

Some Siouan languages make gender distinctions in the choice of imperative marker based on the sex of the speaker (not of the addressee). Biloxi (Siouan (Siouan): USA), a language that went extinct in the 1930s, is claimed to have had a spectacular set of imperative markers, although it may be that the number of markers it is claimed to have had is somewhat exaggerated (see Einaudi 1974: 88ff for a discussion). At the very minimum there were three different imperative markers as well as a zero marker. For plural addressees the additional number marker (*a-*) *-tu* was used.

Biloxi (Siouan (Siouan): USA)

(335) a. *-**ta***: male speaker to male addressee(s)
eyąhį-**ta** (single addressee) **a**-dǫx-tu-**ta** (plural addressees)
come-IMP.M/M PL-look-PL-IMP.M/M
'Come!' 'Look!'

b. *-**di***: male speaker to female addressee(s)
umaki-**di**
bathe-IMP.M/F
'Bathe!'

c. *-**te***: female speaker to male addressee(s)
toho-**te** dǫx-tu-**te**
lie.down-IMP.F/M look-PL-IMP.F/M
'Lie down!' 'Look!'

d. Ø: female speaker to female addressee or male/female speaker to child
handa
stay.IMP.Ø
'Stay here!' (Einaudi 1974: 88ff)

Example (335) shows various forms of imperatives in Biloxi which take different markers depending on the sex of the person issuing the command. Thus *-ta* is used if a male is commanding one or more other males, *-di* is used if a male is commanding one or more females and *-te* is used if a female is commanding one or more males. With plural addressees the plural marker *-tu* appears before the imperative marker (as shown in Example (335a and c); the former also contains the plural prefix *a-*). The zero marker is commonly found and seems to have been used for issuing commands to children and from female speakers to one or more female addressees.[194]

Abkhaz is interesting in that it has an ergative alignment pattern for imperatives. In Example (333) above we have an intransitive verb *c'a* 'go', where the agreement is with the S (the one who is to perform the action). With transitive verbs, however, we get agreement with the P with the prefix *d-* for animates and *y-* for inanimates:

Abkhaz (Northwest Caucasian (Northwest Caucasian): Georgia)

(336) a. y-ga (single addressee)
INANIM$_P$-take
'Take (it)!'

b. d-ga (single addressee)
ANIM$_P$-take
'Take (him/her)!'

c. y-žw-ga (plural addressees)
INANIM$_P$-PL-take
'Take (it)!' (Wolfgang Schulze, p.c.)

In (336a and b) the agreement in the singular imperatives is with the P and not the A, since the agreement marker denotes what kind of animacy the P has, and not, for example, the gender of the person meant to carry out the action. The pattern is therefore an ergative one: the alignment is between S and P. Notice that this does not carry over to the plural, since in the plural the verb is marked for all central arguments: the S with intransitive verbs and both the A and the P with transitive verbs, as shown in (336c). Alignment was discussed in 9.1.3.

In some languages, such as Welsh, a subset of verbs marks the imperative with a suppletive form.

---

194. Baloxi also had gender sensitive markers for other sentence types: for example, declarative clauses carried the optional markers *-na* (if the speaker was male) and *-ni* (if the speaker was female), while interrogative clauses carried the obligatory markers *-wo* (if the speaker was male) and -Ø (if the speaker was female). See further Einaudi (1974:79ff).

Welsh (Indo-European (Celtic): UK)

(337)

	SINGULAR ADDRESSEE	PLURAL ADDRESSEES
mynd 'to go'	dos (N)/cer (S) 'go!'	ewch (N)/cerwch (S) 'go!'

(King 1996: 226)

The verb *mynd* 'go' takes a suppletive form both in the singular and in the plural imperative; the two forms differ between the northern and the southern dialects (N and S respectively). Notice that there is no other number marking for the imperative than the choice of suppletive form. If a language has suppletive imperatives, it is likely that it is for basic motion verbs such as 'come' and 'go' (Veselinova 2006: 139).

Languages may have different imperative markers depending on aspect, remoteness, distance, direction or evidentiality. Lavukaleve, for example, has two sets of imperative markers in three numbers (singular, dual and plural), for the perfective and imperfective aspects respectively.

Lavukaleve (Solomons East Papuan (Lavukaleve): Solomons Islands)

(338)

	PERFECTIVE	IMPERFECTIVE
SINGULAR	-va	-ma
DUAL	-ila	-mela
PLURAL	-iva	-ba

a. iru-**ma**
sleep-IPFV.IMP.SG
'Sleep!'

b. iru-**va**
sleep-PFV.IMP.SG
'Shut your eyes!' (Terrill 1999: 311f)

While some verbs in Lavukaleve are incompatible with only one of the sets of imperative markers due to the actionality of the verb (for instance, *ngoa* 'stay' can only take the imperfective markers while *foa* 'go.down' can only take the perfective markers), others may take either set, which, in turn, affects the translation of the command. Thus *iru* 'sleep' translates into the command 'shut your eyes' when marked with the perfective imperative endings, as shown in (338b).

Some languages have a special 'delayed' or 'remote' imperative (sometimes also called a future imperative), which contrasts with the direct imperative.

Tucano (Tucanoan (Tucanoan): Colombia, Brazil)

(339) a. ba'á-ya
eat-IMP
'Eat!' (now)

b. ba'á-apa
eat-FUT.IMP
'Eat!' (later) (West 1980: 51)

In Tucano the imperative suffix indicates whether the action has to be carried out immediately ((339a)) or only at some point in the future ((339b)). Very often this 'delayed' or future imperative is considered a politer form of command.

Many Cariban languages have different imperatives depending on motion distinctions. Tiriyo, for example, has three different sets of markers for two sets of numbers:

Tiriyo (Cariban (Cariban): Suriname, Brazil)

(340)

	SINGULAR ADDRESSEE	COLLECTIVE ADDRESSEE
STATIVE	-(kë)	-tëkë
ALLATIVE (motion away)	-ta	-tatë
VENITIVE (motion to)	-mïī	-mïīko(mo)

a. apëh-kë
   3O:catch-STATIVE.IMP.SG
   'Catch it!'

b. papa=pa i-wa-ta
   2:father=RPT 3OBJ-fetch-ALL.IMP.SG
   'Go get your father!'

c. epë-e=mïī
   bathe-RPT=VEN.IMP.SG
   'Come bathe!' (Meira 1999: 319ff)

The stative imperative does not imply any motion, while the allative imperative implies motion away from and the venitive implies motion towards the speaker. The allative imperative in Tiriyo is also what may be termed a 'distal imperative': the action has to be carried out elsewhere than where the order was given.

Languages may have different imperative markers depending on the evidentiality of the command. In Tariana (Arawakan (Arawakan): Brazil), for example, an imperative uttered by one person may be repeated by another person, but then only as a second-hand command ((341a)). Also, imperatives of warning ('apprehensives') take a different form depending on whether the addressee can see the danger or not.

Tariana (Arawakan (Arawakan): Brazil)

(341) a. pi-ñha-pida
   2SG-eat-SEC.IMP
   'Eat!' (Eat-(as)-you.were-told-to.)

b. pi-na di-pasya-da
   2SG-OBJ 3SGNF-squash-VIS.APPREH
   '(Beware, a car) might squash you.' (You are walking in the middle of the road and can see it.)

c. pi-wha-ñhina
   2SG.OBJ-fall-APPREH.NONVIS
   'You might fall down.' (You are not looking.) (Aikhenvald 2010: 138, 141)

In (341a) an order previously given by person A is repeated by person B (a second-hand command). In (341b) the addressee is being warned about a danger s/he can see while in (341c) the addressee is being warned about a danger that neither speaker nor addressee can see.

12.1.3.2 *Negative imperatives*

All known languages have a way of commanding someone not to do something, but how this is expressed may differ. **Negative imperatives**, or **prohibitives**, are the device used to tell someone not to carry out an action. An example of a prohibitive is *Don't touch that!* where the addressee is commanded to refrain from doing something (in this case touching the item in question).

Of the 495 languages in van der Auwera & Lejeune's (2011b) database on prohibitives, 113 (or 22.8%) form a prohibitive by way of the positive imperative and the normal negative, that is, there is no special form for the prohibitive. Maithili is an example of such a language.

Maithili (Indo-European (Indic): India, Nepal)

(342) a. j-**o**
go-IMP(2NH)
'Go!'

b. nəi j-**o**
NEG go-IMP(2NH)
'Don't go!' (Yadav 1996: 284, 289)

In (342) the imperative is formed by using a person/number dependent imperative verb ending, while the prohibitive is formed by simply adding the ordinary negator to the imperative clause. English also falls into this category of languages, as shown in the translations above; however, for the prohibitive English requires the auxiliary *do* to be added to the clause.

Somewhat more common than this strategy is to form prohibitives with the positive imperative plus a special negative. This is found in 182 languages (36.8%) in van der Auwera & Lejeune's database. Purépecha is an example of such a language, where the negative for prohibitives is *áši* while the negative for other sentence types is *nó*:

Purépecha (Tarascan (Tarascan): Mexico)

(343) a. á-Ø í-ni awárdienti
drink-IMP.2 DEM-OBJ wine
'Drink this wine!'

b. **áši** xwá-Ø-rini sáni
NEG bring-IMP.2-OBJ few
'Don't bring me few!'
c. xí **nó** tʰirí-ša-ka
1SG NEG eat-PROG-ASS.1
'I am not eating.' (Chamoreau 1998: 59–60)

The imperative marker -Ø- remains the same for both positive and negative imperatives ((343a and b)), but the negation differs between negative imperatives and negative declaratives ((343b and c)).

Another fairly common strategy is to have a special negator for prohibitives and a form of the verb that is different from positive imperatives, found in 145 languages (29.3%) in van der Auwera & Lejeune's sample. In Lango, for example, the prohibitive is formed by using a special negative and the subjunctive mood (not the imperative) for the verb:

Lango (Nilo-Saharan (Nilotic): Uganda)

(344) a. dăc
2SG.drop.IMP
'Drop (it)!'
b. **kŏr** ìdăc
NEG 2SG.drop.SUBJ
'Don't drop (it)!'
c. án **àpé** àwótò Kàmpálà
1SG 1SG.NEG 1SG.go.PFV K.
'I didn't go to Kampala.' (Noonan 1992: 142–3)

In (344a) the clause is in imperative mode, marked with a zero morpheme. In (344b) the clause is a prohibitive and therefore takes the special negative *kŏr* 'not', but the verb is the subjunctive mode rather than the imperative. The negator for other sentence types than prohibitives is *pé* (plus the person/number inflection), as shown in (344c).[195] It should be noted that van der Auwera & Lejeune have included in this category languages with pure prohibitive markers, that is, markers that specifically denote prohibition, which seems to be a rather rare phenomenon. In König & Siemund's (2007) database of 70 languages, only four of them (Lezgian, Macushi, Malayalam and Warekena) have specific prohibitive markers. Kayardild, which is not in König & Siemund's sample, is an example of such a language:

195. It could be argued, however, that the Lango imperative is actually a suffixless subjunctive. See Noonan (1992: 143).

Kayardild (Australian (Tangkic): Australia)

(345) a. duura-**tha** ngad
poke-IMP 1SG.NOM
'Poke me!'

b. duura-**na** ngad
poke-PROHIB[196] 1SG.NOM
'Don't poke me!' (Evans 1995: 256)

As Example (345) shows, the positive imperative in Kayardild is expressed with the suffix *-tha*, while the negative imperative (prohibitive) is expressed with the suffix *-na*.

The fourth and smallest group of languages (55 languages or 11.1%) in van der Auwera & Lejeune's sample form prohibitives by using the same negator as for other sentence types, but a verb form that differs from the positive imperative. Italian is an example of such a language for the singular imperative:

Italian (Indo-European (Romance): Italian)

(346) a. parl-**a**!
speak.2SG.IMP
'Speak!'

b. **non** parl-are!
NEG speak.IND
'Don't speak!'

c. **non** parl-i bene
NEG speak-2SG.PRES well
'You don't speak well.' (source: personal knowledge)

In (346a) the verb takes the imperative personal ending, but in the prohibitive in (346b) the verb is in the indicative. The negator *non* is the same as in a declarative sentence, as shown in (346c).

The figures in WALS are minimally affected once the five contact languages have been subtracted, four of which (Angolar (Creole (Portuguese-lexified): São Tomé & Príncipe), Naga Pidgin (Creole (Assamese-lexified): India), Ndyuka and Sango) form the prohibitive by way of the normal imperative and the normal negative, while the fifth (Ma'a/Mbugu) has a special imperative and a special negative for prohibitives. The pattern for prohibitives differs quite considerably among the APiCS languages, as shown in Table 12.1.

196. This is glossed 'NEGIMP' in the source; for the sake of conformity with the discussion I have altered the gloss to PROHIBitive.

Table 12.1 Comparison between WALS and APiCS languages for prohibitives. Adapted from van der Auwera & Lejeune (2011b) and Michaelis et al. (2013: feature 56). Absolute numbers in parentheses.

	Value	WALS	APiCS[197]
1.	Normal imperative + normal negative	22.85% (113)	67.6% (50)
2.	Normal imperative + special negative	36.8% (182)	18.9% (14)
3.	Special imperative + normal negative	29.3% (145)	9.5% (7)
4.	Special imperative + special negative	11.1% (55)	1.4% (1)
	Total	495	74

It thus seems that that pidgin and creole languages are more likely than non-creole languages to form prohibitives with a positive imperative plus the normal negative.

12.2 Politeness

At all times when people interact, identities and senses of selves are jostling for attention. A central concept in pragmatics is the concept of **face**, that is, the public self-image that we all want to maintain (a term ultimately derived from the English expression 'to lose face', see Brown & Levinson (1987 [1978]: 61). There are two aspects to the concept of face, the **negative face**, which basically denotes the need to be independent and free from imposition, and the **positive face**, which, very simplified, denotes the need to belong and be accepted.

> [F]ace is something that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction. In general, people cooperate (and assume each other's cooperation) in maintaining face in interaction, such cooperation being based on the mutual vulnerability of face […] since people can be expected to defend their faces if threatened, and in defending their own to threaten others' faces. (Brown & Levinson 1987 [1978]: 61)

Politeness is a way of interaction which shows awareness of and respect for someone else's face. A **face threatening act** is an act that threatens the integrity and self-image of another person. For example, giving someone a direct order such as *Sit down and be quiet!* implies having social power over that person. It is acceptable for a parent

197. The language with value 4 is Ma'a/Mbugu. For two APiCS languages there is no dominant strategy: in Chinese Pidgin Russian it is equally common to form the prohibitive with the positive imperative and the ordinary negative as it is to form it with the positive imperative and a special negative. In Pichi it is equally common to form the prohibitive with the positive imperative and the normal negative as it is to form it with a different form of the verb and the normal negative. Data is missing for two languages (Batavia Creole and Yimas-Arafundi Pidgin).

to give his or her child a direct order, but giving a direct order to your colleague or neighbour is in general a face threatening act and therefore not acceptable. Conversely, if we act or behave in a way that lessens a possible threat to another person's face, we engage in a **face saving act**. If, for example, you wish to get on with the meeting and that your colleague would stop roaming about the room talking on his mobile phone, you might convey this by using an indirect speech act like *Right, should we sit down and continue?* where the question removes the assumption of social power (for instance, by implying that the addressee has the theoretical option to answer however he pleases) and therefore maintains the balance of face.

This section will give a very brief sketch of how indirect speech acts differ from direct speech acts, as well as mention some ways in which languages have grammaticalized politeness. It is beyond the scope of this section to give a comprehensive discussion of all the aspects involved in the study of linguistic politeness. For a very accessible introduction to the topic of linguistic politeness, see Watts (2003). A classic piece of work, around which much discussion has centred, is Brown & Levinson (1987 [1978]).

12.2.1 Indirect speech acts

Speech acts may be either direct or indirect. In a **direct speech act** the content of the utterance matches the speaker's intention of the utterance. Thus the declarative is simply a statement uttered with the purpose of giving information, while the interrogative is simply a question uttered with the purpose of getting information and the imperative is simply a command with the purpose of getting someone to do something. In other words, when we engage in direct speech acts, we mean what we say. The sentence types discussed in 12.1 were all direct speech acts. However, as mentioned, social situations involve complex structures of cooperation: we might want to achieve something, but we also have to take into account that we are surrounded by others whom we do not want to offend. As mentioned above, giving someone a direct command, for example, implies that we either have power over or are very intimate with that person (or both). This may or may not be a useful social strategy to engage in, depending on the situation. A parent can pretty safely issue a direct command to his or her child while the child still lives at home. A military commander can pretty safely issue a direct command to a soldier. There might be situations where close friends can give each other direct commands. But it is rarely useful for an employee to give a direct command to his or her boss, or for a host to order his or her guests about (or the guests the host, for that matter). Therefore we engage in various politeness strategies.

One way to save someone's negative face, is, for example, to show concern about imposition. If we want help from somebody, we might phrase the request with an initial face saving phrase, such as *Excuse me, …* or *I'm sorry to bother you, …* or

something similar. A way of saving someone's positive face would be, for example, to increase the sense of group belonging. One common strategy for that in English is to use an inclusive *we*, thereby reducing the request into a common act. An example of this would be the indirect speech act mentioned above (*Right, should we sit down and continue?*), where the pronoun *we* gives a tone of group belonging, even though there is only one intended addressee. There's a fine balance to these kinds of strategies though: an inclusive *we* can very easily sound ironic and get a superior tone, thereby becoming a face threatening rather than face saving device.[198]

A very common face saving device is to use the aforementioned indirect speech acts to achieve our purposes. In an **indirect speech act** the content of the utterance does not correspond to the speaker's intention of the utterance. We may, for example, utter a question or a statement, but the real purpose of the utterance is to get someone to do something for us (i.e. it is an indirect command). In other words, when we engage in indirect speech acts, we actually do not mean exactly what we say. For instance, if a stranger carrying a pile of books asks you *Do you know where the library is?* you are likely to interpret the question as a command (*Tell me where the library is*) and answer it by giving directions on how to get there, rather than take it as a direct interrogative speech act and answer something like *Yes, thank you, I know where it is*. Likewise, if you and your friend are about to get into a car and the friend utters the declarative sentence *The door on this side is locked*, you are likely to interpret that as an indirect command (*Open the door for me*), rather than just file it as a piece of information and drive off. It is very common for speakers to use declaratives and interrogatives as politely disguised requests (or commands).

There is always room for misunderstanding between individuals and cultures when it comes to interpreting speech acts. What one person might find a clear case of an indirect command, the other might find a clear case of a direct speech act. What is polite for one person might not be for the other, and so on. For instance, the individual who has been raised to never directly ask for something might, when having dinner with colleagues, ask her neighbour *Would you like some more potatoes?* in the expectation that the neighbour will then ask her the same question (to which she can then safely say yes). If the neighbour, however, does not consider it impolite to ask for something, s/he might not understand the question as an indirect speech act, and might therefore simply answer *No, thank you*, in which case there has been a miscommunication. This is of course especially true for clashes between people of different cultures.

---

198. The so-called 'medical *we*' is especially famous for this, where a nurse or a doctor uses the first person plural pronoun when actually addressing the patient, such as *Have we taken our medicines, then?* or *How are we today?* which is actually in many situations considered patronizing. See, for example, Wales (1996: 63ff).

12.2.2 Honorifics

Many languages grammaticalize various politeness strategies (for more on grammaticalization, see 13.1). **Honorifics** can, very simplified, be thought of as linguistic politeness devices, or devices for marking social distinction (so-called 'social deixis'). Following Brown & Levinson (1987 [1978]) and Levinson (1983) I will distinguish between three types of honorifics depending on the relationship between the speaker and others: the referent honorifics, which have to do with the relationship between the speaker and things or persons referred to; addressee honorifics, which have to do with the relationship between the speaker and the addressee; and bystander honorifics, which have to do with the relationship between the speaker and the 'bystanders' or overhearers (covering such things as audiences as well as bystanders who are not participants of the interaction).[199]

12.2.2.1 *Referent honorifics*

With a **referent honorific** the politeness or respect distinctions are conveyed by referring to the target of the respect, that is, the choice of the linguistic form is dependent on what is referred to. A well-known instance of referent honorifics is the choice of pronoun when addressing a person. In European languages it is very common to have a binary distinction of politeness, where one form is considered familiar and the other polite, as in the French distinction *tu*/*vous*, both of which refer to a single person. This binary distinction in pronouns of address is often called **T/V pronouns**, after the Latin *tu* 'you.SG.NOM' and *vos* 'you.PL.NOM'. In his sample of 207 languages on politeness distinctions in second person pronouns, Helmbrecht (2011) found that 49 languages (23.7%) make use of this kind of binary distinction, many of which cluster in Europe. It should be noted here, however, that this figure comprises languages with vastly different levels of T/V pronoun usage. In German, for example, the polite form *Sie* is the standard form of address between adults who are not in a close social relation. In Swedish, however, which is also listed as having a binary politeness distinction, this is not the case. While it is imaginable to use the plural pronoun form *ni* to address a single person, the context would be highly marked indeed. The default mode of address between adults who don't know each other is with the singular ('informal') *du*. Having Swedish as one of my mother tongues, I would, in fact, be hard put to find a situation where using the plural ('formal') *ni* to address a single person

199. Brown & Levinson (1987 [1978]) and Levinson (1983) have a fourth social deixis axis, that of the setting, where the formality of the language the speaker uses is dependent on the setting where the utterance takes place. As this concerns formality levels and not honorifics, it will not be dealt with further here.

would feel adequate and that would not convey an awkward, almost ironic tone. The usage of universal *du* in Swedish is a rather recent phenomenon and, interestingly enough, it seems that the language might be changing back to making politeness distinctions. The generation born in the later 1980s, and even more so those born in the 1990s, find *ni*-usage more acceptable than those of us born in the 1970s or before (Charlotta Busing, p.c.).

Some languages, 15 (7.2%) in Helmbrecht's database, mainly clustered in South Asia but also found elsewhere, use multiple politeness distinctions for the second person pronoun. Nepali is an example of a language with three levels of politeness distinctions in the singular and two in the plural:

Nepali (Indo-European (Indic): Nepali)

(347)

	SINGULAR	PLURAL
LOW GRADE HONORIFIC ('informal')	tã	timī-haru
MID GRADE HONORIFIC ('polite')	timī	timī-haru
HIGH GRADE HONORIFIC ('superpolite')	tapāï	tapāï-haru

(Acharya 1991: 106)

In (347) we have three different forms for the second person pronoun in the singular depending on the level of politeness, while the plural adopts the two politer forms and adds a plural marker to them.

There are also languages where a pronoun is avoided for the sake of politeness, a strategy sometimes termed '**pronoun avoidance**'. In these languages it may be considered face threatening to directly address a person, so other kinds of terms are used instead, such as status and kinship terms or various kinds of titles, and so on. Seven languages (3.4%) in Helmbrecht's database make use of this kind of strategy, all of them in East and Southeast Asia. In Indonesian, for example, kinship terms such as *saudara* 'sibling; relative of same generation' function as an impersonal pronoun used between speakers of the same generation (or by a speaker to somebody younger) when they are not well acquainted, while such terms as *bapak* 'father' or *ibu* 'mother' are used as a respectful address to people older than the speaker and between adults of marriable age. Furthermore, very often people will avoid using the pronoun even when on an equal status level, and instead use the addressee's name (effectively saying something like 'Does Tom want more tea?' when addressing Tom) (Sneddon 1996: 161ff; see also Wallace 1983 for a discussion on the origins of Indonesian pronouns and what that can tell us about cultural attitudes and ideology).

The majority of the languages in Helmbrecht's database (136 of 207 or 65.7%), however, do not make any politeness distinctions for the second person pronominals. The languages in APiCS display an almost identical pattern, as shown in Table 12.2.

Table 12.2 Comparison between WALS and APiCS languages for politeness distinction in pronouns. Adapted from Helmbrecht (2011) and Michaelis et al. (2013: feature 18). Absolute numbers in parentheses.

	Value	WALS[200]		APiCS[201]	
1.	No politeness distinction	66.4%	(136)	63.5%	(47)
2.	Binary politeness distinction	22.9%	(47)	24.3%	(18)
3.	Multiple politeness distinction	7.3%	(15)	5.4%	(4)
4.	Pronoun avoidance	3.4%	(7)	6.8%	(5)
	Total		205		74

Pidgin and creole languages thus do not seem to behave differently from non-creoles with respect to politeness distinctions in pronouns.

Notice that with referent honorifics, which pronoun politeness is a part of, the choice of form is dependent on who or what is being referred to. With the second person pronouns the referent and the target happen to be the same. But we may also have honorific distinctions in the third person. In Korean, for example, the choice of the third person pronoun is dependent on what is being referred to and what level of politeness is required:

Korean (Isolate: N, S Korea)

(348)	3RD PERSON	SINGULAR	PLURAL	
	THING	D-kes 'it'	D-kes-tul 'they'	
	CHILD	D-ay 's/he'	D-ay-tul 'they'	
	ADULT: FAMILIAR	D-salam 's/he'	D-salam-tul 'they'	
	ADULT: BLUNT	D-i 's/he'	D-i-tul 'they'	
	ADULT: POLITE	D-pun 's/he'	D-pun-tul 'they'	(Sohn 2001: 207)

In (348) we have the third person pronoun singular and plural, which consists of the definite demonstrative form (noted as D here) plus the relevant suffixes. While Korean makes an animate/inanimate distinction, just as English does (*it* versus *s/he*), the choice of the suffixes for the animate referents are not dependent on gender, as in English, but on social deixis, or politeness. Thus a child is referred to with the demonstrative plus the suffix *-ay*, while an adult is referred to with the demonstrative

200. For the purpose of comparison between APiCS and WALS patterns, the creole languages in the WALS sample have been taken out, which is why the figure for WALS differ slightly here from those earlier in the text.

201. The languages for value 3 are Ambon Malay, Korlai, Zamboanga Chabacano and Caviteño Chabacano. The languages for value 4 are Afrikaans, Diu Indo-Portuguese (Creole (Portuguese-lexified): India), Papiamentu and Principense (Creole (Portuguese-lexified): São Tomé & Principe). Data is missing for Batavia Creole and Juba Arabic (Pidgin (Arabic-lexified): Sudan).

plus one of three available suffixes (*-salam*, *-i*, and *-pun*). The choice of suffix for adult referents depends on the level of politeness required. Compare this to English, where a child and an adult will be referred to by the same pronoun (*he* or *she*) and where there are no honorific distinctions at all (a family member will be referred to with the same pronoun as a member of the Royal House of England). Notice, further, that English makes no distinctions whatsoever in the plural (everything translates into 'they').

12.2.2.2 *Addressee honorifics*

An **addressee honorific** system is one where the choice of the linguistic form is dependent on the addressee of the utterance. Virtually any part of the linguistic system may be affected by addressee honorific requirements. We have already seen that languages may have morphologically variable forms depending on levels of politeness. This is also found in verb forms, such as different kinds of imperatives carrying different kinds of politeness connotations. We have seen that languages may make a distinction between immediate and distal imperatives, where the immediate imperative is a familiar (or maybe even rude) command while the distal imperative is a polite command.

Epena Pedee (Choco (Choco): Colombia)

(349) a. pʰáta kʰó-tɨ́
plantain eat-IMP.PL
'Eat the plantain!'

b. tʰɨ́pɨ́ pʰua-pʰéda a-hi
firewood blow-POL.IMP say-PAST
'He said, "you will have to fan the fire".' (Harms 1994: 129f)

In Epena Pedee the polite imperative, which receives its own morphological marking (compare *-ti* in (349a) with *-pʰéda* in (349b)), indicates that the action is going to have to be carried out at some point in the future. The distal nature of this kind of imperative makes it less direct and therefore, logically, more polite.

Korean verbs are not only marked with TMA and sentence type affixes but also a host of affixes relating to addressee honorifics. A simple statement like *It is raining* thus takes six different forms depending on the social relationship between the speaker and the addressee:

Korean (Isolate: N, S Korea)

(350)

plain	pi	ka	o-n-ta
	rain	NM	come-IN-DC
intimate	pi	ka	w-a
	rain	NM	come-INT
familiar	pi	ka	o-ney
	rain	NM	come-FML

blunt	pi	ka	o-o
	rain	NM	come-BLN
polite	pi	ka	w-a.yo
	rain	NM	come-POL
deferential	pi	ka	o-p-ni-ta
	rain	NM	come-AH-IN-DC
neutral	pi	ka	o-t-a
	rain	NM	come-IN-DC
	'It is raining.'		(Sohn 2001: 269)

Example (350) shows the six levels of politeness in Korean. The choice of level depends on the level of intimacy between speaker and addressee. The PLAIN level is used with children, or to the speaker's younger siblings, children and grandchildren (regardless of age), to the speaker's daughter-in-law and between adult friends who have been friends since childhood. The INTIMATE level (also called 'half-talk style') is used by small children to any of their family members (including elders), to young students, to the speaker's son-in-law, and between adults whose friendship began when they were young (children or teenagers). The FAMILIAR level is a bit more formal than the intimate level and is used by male speakers to young adolescents or young adults (e.g. college students), to the speaker's son-in-law, or between close adults whose friendship began in adolescence. The BLUNT level is authorative, such as utterances spoken by bosses to subordinates. This level is gradually disappearing. The POLITE level is the most widely used level and is basically the informal counterpart of the deferential level. The polite level is used between close adults, but also in daily conversations between distant adults or to superiors. The DEFERENTIAL level is a more formal counterpart of the polite level. It is used between distant adults or to superiors, or in such contexts as verbal news reports and formal announcements. The NEUTRAL level looks like the plain level except with interrogatives and imperatives. It is used in quotatives, in writing to general audiences (i.e. in books, articles and newspapers), and in written exam instructions. There is also a seventh politeness level, the SUPERPOLITE level, which is now obsolete, but can be found in religious prayers, poems and occasionally in extremely formal or deferential letters. See further Sohn (2001: 271ff).

'HUMBLE WE'

Rather than a 'royal we' (*pluralis majestatis* or 'majestic plural'), where the first person plural *we* is used to refer to the first person singular *I*, some languages employ a 'humble we': in Sawu (Austronesian (Central Malayo-Polynesian): Indonesia) the first person plural inclusive pronominal *dii* 'we' may be used as a polite or deferential form for 'I' (first person singular) (Cysouw 2005: 218).

Some languages, such as Javanese, have a so-called 'honorific register', where the choice of vocabulary is addressee honorific, i.e. the choice of which words to use for an utterance is dependent on the relationship between the speaker and the addressee. There are three main speech levels: *ngoko* (informal), used only when speaking to someone very familiar, *madyô* (semi-polite), used with people of an intermediate level of familiarity, such as neighbours, and *krômô* (polite), used with distant adults. A fourth type of vocabulary, a kind of "respect vocabulary", may be used in connection with any of the other level types to convey an added level of respect. This fourth type, the "respect vocabulary" has two subcategories, *krômô inggél* (lit. 'high krômô'), used for highly respected persons, and *krômô anḍap* (lit. 'humble krômô'), "used in referring to any person's actions towards a highly respected person" (Poedjosoedarmo 1968: 58). We may thus end up with five different words that translate the same way into English, but that carry different levels of politeness.

Javanese (Austronesian (Javanese): Indonesia)

(351) NGOKO MADYÔ KRÔMÔ KRÔMÔ' INGGÉL KRÔMÔ ANḌAP
akôn kèn kèngkèn/puréh ḍawoh ng-atur-i
'tell/ask to do something' (Poedjosoedarmo 1968: 58)

In (351) there are five different words that all translate into the same thing in English, but the choice of which is dependent on the level of respect the speaker wishes or is obliged to show the addressee. There are then further distinctions in the choice of grammatical affixes. The combination of vocabulary and affixes leads to nine different levels of politeness. The three main speech level vocabularies have different sizes: *ngoko* is the basic vocabulary which carries every kind of concept, *krômô* has about 850 words, *krômô inggél* has about 260 words, *madyô* has about 35 words and *krômô anḍap* has about 20 words (Myhill 1994: 75f).

12.2.2.3 *Bystander honorifics*

With **bystander honorifics** the linguistic form of the language is not dependent on the speaker or on the addressee, but on who is within earshot of the utterance. That is, it is not dependent on the relationship between the speaker and the addressee, nor is it dependent on what is being referred to, but simply who can hear what is being said. This therefore covers participants, such as audiences, as well as non-participants, or 'bystanders'. This is often termed 'avoidance language' or 'honorific register'. Many Australian languages had or have bystander honorifics to varying degrees. Dyirbal is famous for having had two language variants, Guwal and Dyalŋuy. Guwal was used in all circumstances except when certain 'taboo relatives' were present, in which case Dyalŋuy had to be used.

> No man or woman would closely approach or look at a taboo relative, still less speak directly to them. The avoidance language, Dyalŋuy, had to be used whenever a taboo relative was within earshot. The taboo was symmetrical – if X was taboo to Y so was Y to X.
>
> Taboo relatives were:
>
> [1] a parent-in-law of the opposite sex; and, by the symmetry rule, a child-in-law of the opposite sex.
> [2] a cross-cousin of the opposite sex – that is, father's sister's or mother's brother's child. (Dixon 1972: 32)

Guwal and Dyalŋuy were identical phonologically and almost identical grammatically, but differed completely in their vocabulary. To give an example:

Dyirbal (Australian (Pama-Nyungan): Australia)

(352)

Guwal	**Dyalŋuy**	
yanu(l)	bawalbil	'go'
buɽal	ŋuɽimal	'see, look at'
ŋalŋga	ŋalmaru	'child'
ŋinay	maḏirabil	'sit, stay, camp'

(Dixon 1972: 314)

While a number of languages may have 'taboo relatives', these are not necessarily identical cross-culturally. Guugu Yimidhirr (Australian (Pama-Nyungan): Australia) also has an avoidance language that specifically involves the vocabulary and prosody of the language. Here the taboo is especially strong between a man and his in-laws.

> a man could not speak at all to his mother-in-law, remaining silent in her presence and absenting himself when possible. With his father-in-law, his brothers-in-law and with certain other relatives, a man was obliged to speak in a specially slow, soft, and respectful tone of voice, and to substitute respectful equivalents for many common words. (Haviland 1979a: 32)

It should be noted that the overlap between the two vocabularies was not complete; the basic, informal 'everyday lexicon' had more items than the 'respectful lexicon'. It was or is, however, still possible to express the same things in both varieties. This was achieved by different methods of circumlocutions. In Guugu Yimidhirr, for example, several 'everyday language' words translate into only one 'respectful language' (or 'brother-in-law language', as it was called) word:

Guugu Yimidhirr (Australian (Pama-Nyungan): Australia)

(353)

EVERYDAY LANGUAGE		BROTHER-IN-LAW LANGUAGE
dhadaa	'go'	balil 'go'
dharmbil	'float, sail, drift'	
yaalgal	'limp'	
gaynydyarr	'crawl'	
biilil	'paddle (canoe)'	
…		

(Haviland 1979b: 371)

In (353) several of the 'everyday lexicon' words translate into the same 'respectful lexicon' word. However, in both Guugu Yimidhirr and in Dyirbal, and presumably in most languages that had or have an 'avoidance language', there was no limit to what could be expressed in the 'respectful language'. In Guugu Yimidhirr, for example, various circumlocutions would be used to achieve the exact expression intended. For instance, if someone specifically wanted to say 'float' in the respectful language, this was done by the expression *balil wabiirrbi* (lit. go on water) and if someone wanted to say 'limp' in the respectful language, the expression would be *dyirrun balil* (lit. go badly), and so on (Haviland 1979b).

Bystander honorifics do not have to be restricted to the lexicon. Waray (Australian (Waray): Australia) also had an avoidance language, referred to as 'sideways language', which was used with various in-law relatives and which involved adding a suffix *-lawu* to nominals and using the plural forms for verbs. There is no record of the avoidance language involving any vocabulary suppletion (Harvey n.y.: 16).

Waray (Australian (Waray): Australia)

(354) nguk-lawu ban-ba-wu gan-a-ga-ng-u
tobacco-AVOID 1SG.OBJ-2PL.S-give IRR-2PL.S-have-NPST-DAT
'Would you(PL) give me any tobacco if you(PL) have any? (addressed to one person)'
(Harvey n.y.: 16)

In (354) the 'sideways language' demands the suffix *-lawu* on the noun (*nguk* 'tobacco') and that the verb is inflected in the plural even though the utterance only has a single addressee. For more on avoidance from an anthropological point of view, see, for example, Stasch (2003) and Pans (1998) with references.

12.3 Speech acts in sign languages

Like spoken languages, all known sign languages have ways to carry out the basic functions of giving information, gleaning information and issuing commands. The declarative is typically the basic, unmarked sentence type. Negative declaratives are typically made in two different ways: with negative particles, which are independent signs, or with a negative intonation, which involves non-manual negative markers that co-occur suprasegmentally with the sentence being signed. For a very thorough discussion on negative constructions in sign languages, see Zeshan (2004a), which I am relying heavily on.

Non-manual negatives are either formed through head movement or facial expressions. There are three main negative marking head movements cross-linguistically. The most common, maybe even universal, negative marking head movement across sign languages is a repeated side-to-side head shake. This non-manual marker

is found in all sign languages in Zeshan's (2004a) sample of 38 languages. The side-to-side head shake may combine with other non-manual markers, such as facial expressions, and may easily combine with manual signs. The second strategy, found in seven of the languages in Zeshan's sample, is to have a single side-ward head turn, where the head is turned to one side and is then kept in that position for the duration of the negation (although the syntactic distribution of this non-manual is more limited than the side-to-side head shake). The third strategy involves a backward tilt of the head, where the head is tilted backwards and remains in that position for the duration of the negation. This is found in ENG, TİD and Lebanese Sign Language (LIB: Lebanon).[202]

Non-manual facial negators tend to be less grammaticalized and therefore less systematic. They usually involve "negative" mimics, such as manipulating the eyebrows (lowering or frowning, etc.), the eyes (narrowing or squinting, etc.), the mouth (drawn down, pursed or spreading the lips, etc.) and the nose (e.g. wrinkling). Non-manual facial negators tend to co-occur with the negative marking head movement. In fact, in ST a non-manual facial negator without an accompanying negative head movement would be interpreted as an expression of emotion or attitude, rather than a grammatical negation (Ahlgren & Bergman 2006).

It is very common for sign languages to have a negative particle, that is, a separate sign that functions as a negator. Neither the negative particles nor the non-manual negative markers shape the signs being negated. That is, the negative particle or non-manuals are simply added; the other signs remain the same as in an affirmative declarative. Many sign languages also have so-called irregular negatives, where the signs in the negative are not the same as in the affirmative (effectively suppletion). For instance, the LIB signs for KNOW and NOT.KNOW look entirely different: KNOW is formed by pointing the index finger to the side of the head (neutral expression), while NOT.KNOW is formed by moving the flat open hand (palm in) from the mouth towards the addressee (plus a non-manual facial negator), see Figure 12.1.

Irregular negatives typically involve signs of cognition (KNOW/NOT.KNOW, UNDERSTAND/NOT.UNDERSTAND, etc.), emotional attitude (WANT/NOT.WANT, LIKE/NOT.LIKE, CARE/NOT.CARE, etc.), modality (CAN/NOT.CAN, MUST/NOT.MUST, NEED/NOT.NEED, etc.), possession and existentials (HAVE/NOT.HAVE, EXIST/NOT.EXIST, GET/NOT.GET, etc.), tense and aspect (WILL/WILL.NOT, DID/DID.NOT, FINISHED/NOT.FINISHED, etc.) and evaluation (RIGHT/NOT.RIGHT, POSSIBLE/NOT.POSSIBLE, ENOUGH/NOT.ENOUGH, etc.) (Zeshan 2011a).

---

202. A point worth noting is that the hearing population of the Eastern Mediterranean, as well as Bulgaria, also use this as a negative gesture. However, as Zeshan (2004a: 11) points out, this negative head gesture is found among hearing people in areas where the sign languages don't have it, such as among the hearing in Italy and Israel (but LIS and ISL do not have this negation head movement), and is among the hearing population often accompanied by a click sound.

Figure 12.1 The LIB signs for 'know' and 'not.know' (Zeshan 2004a: 43).
Used with permission.

Most sign languages have several such irregular negatives. Of the 35 languages in Zeshan's (2011a) database on irregular negatives in sign languages, 21 (60%) have more than five such irregular negatives. BSL, for instance, has at least eight:

BSL (Sign Language: UK)

(355)	evaluation:	NOT.GOOD	
	cognition:	NOT.KNOW	
	emotional attitude:	NOT.WANT; NOT.LIKE; NOT.AGREE; NOT.BELIEVE	
	possession:	NOT.HAVE	
	tense:	WILL.NOT	(Deuchar 1987: 37)

Ten languages (28.6%) in Zeshan's database have 2–5 irregular negatives. LIS, for example, has four. Three languages (8.6%), the Indian dialect of IPSL, KK and LSE, have only one irregular negative, while only one language (2.9%), the Karachi dialect of IPSL, has none.

	LIS (Sign Language: Italy)		**KK** (Sign Language: Indonesia)	
(356)	modality:	CANNOT	aspect:	NOT.YET
	evaluation:	NOT.POSSIBLE		
	emotional attitude:	NOT.WANT; NOT.LIKE		
	(Brunelli 2006: 56)			(Zeshan 2011a: 561)

Sign languages also have ways of forming polar questions and content questions. All 35 sign languages in Zeshan's (2004b) database make use of non-manual signs for polar questions. Typical non-manuals for polar questions are: raised eyebrows, wide open eyes, eye contact with the addressee, head forward position and body forward posture. Contrary to many spoken languages, syntactic changes to the clause (such

as change of word order) are not obligatory in the sign languages in Zeshan's database. There is, however, a tendency for questions containing pronouns to either move the pronoun to sentence final position or to repeat the pronoun sentence finally, as in TSL:

TSL (Sign Language: Thailand)

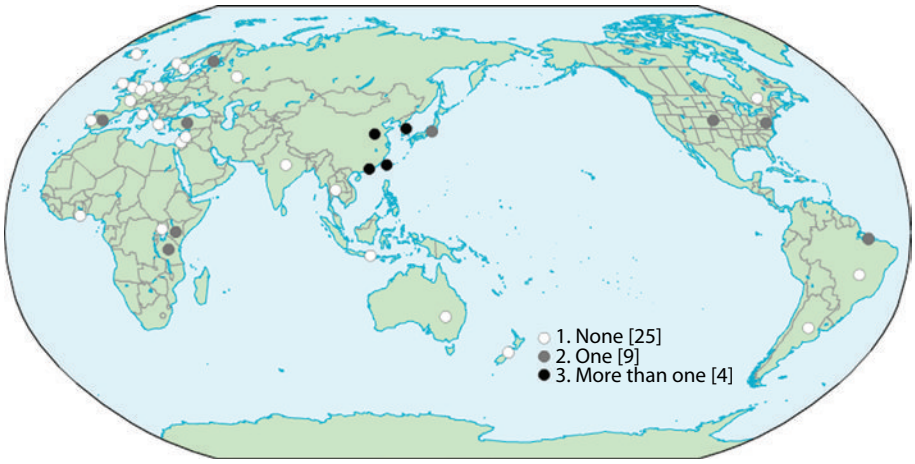
(357) a. _____pol.q
SMOKE INDEX$_2$
'Do you smoke?'

b. _____pol.q
INDEX$_2$ DEAF INDEX$_2$
'Are you deaf?' (Zeshan 2004b: 21f)

In (357a) the personal pronoun (signed by pointing at the addressee) is moved to sentence final position, while in (357b) the pronoun is repeated sentence finally.

Polar questions may also be formed with polar question particles; however, these particles are always optional and always co-occur with the non-manual signs for polar questions. Question particles tend to be used only for polar questions, but a few sign languages, such as SV, also allow them for content questions. There are no known sign languages where the question particle is used for content questions only, or that have different particles for polar questions and content questions (Zeshan 2011b). In her database of 38 languages, Zeshan found that while the absolute majority of languages (25 or 65.8%) lack question particles, about one third have at least one. Nine languages (23.7%) have only one question particle, but four (10.5%), CSL, HKSL, SKSL and TZS, have more than one. While the languages with more than one question particle are all clustered in East Asia, those with one particle are spread over the world, as Map 12.1 shows.

Sign languages form content questions by using a question word and non-manual question signing. The non-manuals for content questions typically, but not universally, differ from the non-manuals for polar questions. The smallest paradigm of question words is to have only one, basically a general WHAT/WHO sign, as in IPSL (Zeshan 2003c: 201). Other sign languages have several question words, such as ISL with six basic question words (WHO, WHAT, HOW.MANY, WHERE, FROM. WHERE/WHO, WHAT'S.THIS) (Meir 2004: 99). Question words may appear either sentence initially, sentence finally, or doubled (i.e. both at the beginning and the end of the sentence). Many sign languages also allow content questions without any question words, a phenomenon not found in spoken languages. This is typically done "by facial expressions (non-manual marking) or by mouthing an imitation of the mouth movements of a corresponding word from the spoken language" (Zeshan 2004b: 30).



Map 12.1 Question particles in sign languages (Zeshan 2011b). For a full legend, see http://wals.info/feature/140A.

As mentioned in 8.6, sign languages tend to form imperatives either with particles or morphological marking. Negative imperatives (prohibitives) are often formed with irregular negation, that is, there is a special sign DO.NOT! marking prohibitives. This contrasts to spoken languages, where pure prohibitive markers are relatively rare.

Politeness and face plays a role in sign languages just as in spoken languages. For instance, indirect speech acts may be used to express wishes and commands, just as in spoken languages. Various linguistic devices also occur. Examples of referent honorifics can be found in the ASL pronominal system, where referencing to a person (first, second or third) by 'pointing' with an open hand (with the fingers together and thumb out) is a polite equivalent of the neutral pronominal signed by pointing with the index finger (Baker-Shenk & Cokely 1991: 207). Another example of a referent honorific is the honorific classifier in TİD. The classifier functions as the typical verbal classifiers of sign languages, modifying the event sign according to the entity engaged in the event. What is unique about this particular classifier is that it is specifically used as a referent honorific. Thus the sign PERSON:COMING is signed with a raised index classifier in neutral, non-honorific, contexts, but with a raised thumb classifier in honorific contexts (Zeshan 2003a: 65). This has so far not been documented for any other sign language.

Addressee honorifics also occur in sign languages, for instance with imperatives. IPSL, for example, has a neutral imperative particle, KARO, and an impolite imperative particle JA:O, which demands immediate action (Zeshan 2003c).

12.4 Summary

All languages are capable of fulfilling three basic functions: to give information, to request information and to issue commands. Any use of these three basic functions in interpersonal communication constitutes a speech act. Speech acts may be direct or indirect. In direct speech acts the meaning of the utterance matches the speaker's intention of the utterance. In indirect speech acts the semantic structure of the utterance does not correspond to the speaker's intention of the utterance.

There are three basic sentence types, declaratives, interrogatives and imperatives. Affirmative declaratives are basic statements. Negative declaratives negate the truth value of a proposition.

Interrogatives are sentence types formed in order to glean information from the addressee and may be either polar questions or content questions. Polar questions are questions to which only a limited number of answers are acceptable, typically either *Yes* or *No*. Content questions are questions which demand answers giving more specific information.

Imperatives are commands. Positive imperatives are formed with the purpose of getting the addressee to do something, while negative imperatives, or prohibitives, are formed with the purpose of getting the addressee to refrain from doing something.

Languages have different ways to uphold various politeness requirements. In most languages indirect speech acts serve as a polite way of issuing commands. Many languages also grammaticalize politeness into various honorifics. Referent honorifics code the relationship between the speaker and whatever or whoever is being referred to. Addressee honorifics code the relationship between the speaker and the addressee. Bystander honorifics code the type of linguistic behaviour a speaker has to engage in depending on who is within earshot of the utterance.

Sign languages also have all three basic sentence types. Negative sentences are formed primarily through non-manual head movement and often also with a co-occurring negative particle. Most sign languages also have lexicalized irregular negatives. Interrogatives are typically formed with non-manual facial signs. Polar questions may have a co-occurring question particle and content questions typically have co-occurring questions words, although content questions without question words are also allowed.

Sign languages also have politeness strategies. Indirect speech acts may be employed to form polite requests, but morphological politeness distinctions can also be found, notably in referent and addressee honorifics.

12.5 Keywords

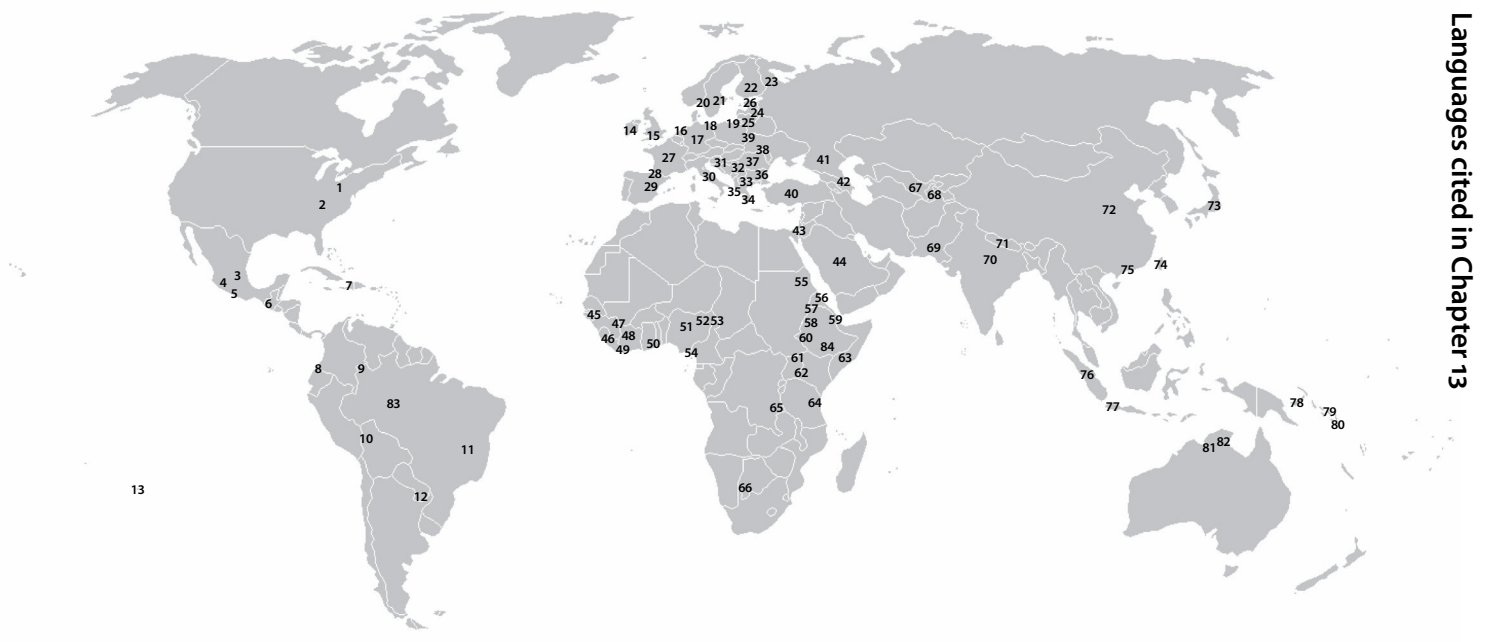
content questions
declarative
direct / indirect speech acts
face
face saving speech acts
face threatening speech acts
honorifics
imperative
interrogative
polar questions
politeness
speech act

12.6 Exercises

1. What are the different strategies for forming prohibitives and how do they pattern in sign languages, non-creole and pidgin and creole spoken languages respectively?
2. How do sign languages pattern with respect to negative declaratives and interrogatives and how does that compare with spoken languages?
3. What is a **face saving** speech act?
4. What is the difference between **referent**, **addressee** and **bystander honorifics**?
5. Is the following statement true or false? Motivate your answer.

   There is no significant difference between pidgins and creoles and non-creole languages with respect to politeness distinctions in pronouns.

Languages cited in Chapter 13



1 German, Pennsylvania
2 American Sign Language
3 Otomi, Mezquital
4 Purépecha
5 Cuitlatec
6 Quiché
7 Haitian Creole
8 Quechua, Imbabura
9 Hup
Tariana
Tucano
10 Mosetén
11 Portuguese, Brazilian
12 Guaraní
13 Rapanui
14 Irish Sign Language
15 British Sign Language
English
English, Middle
English, Old
Welsh
16 Dutch Sign Language
17 German
German Sign Language
18 Proto-Germanic
19 German, Baltic
20 Old Norse
Proto-Norse
21 Swedish
Swedish, Old
22 Finnish
23 Karelian
24 Latvian
Liv
25 Karaim
Lithuanian
26 Estonian
27 French
Old French Sign Language
28 Basque
29 Spanish
Spanish, Middle
Spanish, Old
30 Italian Sign Language
Latin
31 Croatian
32 Serbian
33 Macedonian
Romani, Bugurdži
34 Greek
35 Albanian
Turkish, Macedonian
36 Bulgarian
37 Romanian
38 Rumungro
39 Yiddish
40 Turkish
41 Proto-Indo-European
42 Lezgian
43 Domari
Israeli Sign Language
44 Arabic, Standard
45 Mandinka
46 Kono
47 Maninka
48 Wan
49 Klao
50 Ewe
Likpe
Twi
51 Mopun
52 Kupto
53 Proto-Chadic
54 Ùsàk Èdèt
55 Beja
56 Tigre
57 Ge'ez
Tigrinya
58 Amharic
Awngi
Gumuz
59 Qafar
60 Anywa
Kefa
Wolaytta
Yemsa
61 Acholi
Karimojong
62 Luo
Teso
63 Somali
64 Swahili
65 Swahili, Katanga
66 !Xóõ
67 Uzbek
68 Tajik
69 Indo-Pakistani Sign Language, Karachi
70 Indo-Pakistani Sign Language, Delhi
Sanskrit
71 Manange
Nepali
72 Chinese Sign Language
Mandarin
73 Japanese
74 Taiwanese
75 Cantonese
76 Malay
77 Indonesian
78 Proto-Oceanic
79 Toqabaqita
80 Sa'a
81 Jaminjung
82 Kriol
83 Pirahã
84 Sidaama

Chapter 13

Language change

All languages change, either by way of internal or external factors or both. This chapter will bring up a major factor of internal language change, grammaticalization (13.1). I will give a brief overview of the processes involved in grammaticalization (13.1.1) and the typical paths that have been identified for grammaticalization (13.1.2). I will also very briefly mention the notion of unidirectionality (13.1.3). The principle external cause for language change is contact between languages (13.2). I will first sketch the sociolinguistic factors and mechanisms involved in contact-induced change (13.2.1), before giving a short overview of what kinds of features tend to diffuse (13.2.2). In 13.2.3 I bring up the concept of linguistic areas and give some examples of the most well-known ones. Section 13.3 gives a summary of some grammaticalization paths and contact-induced changes in sign languages.

13.1 Grammaticalization

There is one thing we know with absolute certainty: languages change. All languages are constantly changing and any given description is, in a sense, merely a snapshot of the language at the time of description. While different aspects of the linguistic system in a language may change at different rates, they will all, over time, have changed one way or another. The English that we speak and use today is different from that in Shakespeare's works, and is even more different from that in the Old English epic *Beowulf*, and will be different from that of coming generations. Some features of the English of *Beowulf* that, at the time, were obligatory, have been lost, such as various verbal inflections (e.g. most of the person markings) and nominal inflections (e.g. case marking). Some features of the English of today were not present at the time of *Beowulf*, such as various types of TMA markings (e.g. *be going to* as a future tense marker). Yet other features that are not part of Standard English today but that can be widely found in colloquial Englishes, might enter the linguistic system of Standard English in the future and even become part of a paradigm of obligatory grammatical features. For example, the pronunciation of *be going to* as a future marker may differ from the pronunciation of *be going to* as a lexical verb: compare *I'm gonna hang the laundry now* (which means *I will now hang the laundry* and not **I will now go=propel myself by way of motion to hang the laundry*) with *I'm going to the park now* (which

means *I mean to go=propel myself by way of motion to the park now*). While *going to* may be used both as a future marker (as in *I'm going to hang the laundry now*) as well as a lexical verb of motion, the reduced form *gonna* cannot be used as a verb of motion: something like **I'm gonna the park now* is not acceptable. In other words, *be going to* as an auxiliary has, in colloquial English, acquired a different form (*gonna*) distinguishable from *be going to* as a lexical verb. It may be that this differentiation will one day become obligatory in Standard English, to the extent that even the written language may have an obligatory distinction between the two.

The process by which grammatical forms emerge and develop in languages is called **grammaticalization**. This is essentially a process by which less grammatical forms become more grammatical, for instance lexical items become grammatical items and grammatical items become even more grammatical. At the same time various morphosyntactic and phonological processes tend to occur, such as a reduction in the form of the unit (compare the reduced *gonna* with the original *going to*). The full cycle of grammaticalization has been reached when the form is ultimately lost and new forms take its place. Very simplified, this could be argued to be a cycle where constructions become more synthetic as forms fuse, or more analytic as fused forms are lost and are replaced by lexical items that have developed into new grammatical items.

This section will first give a very brief overview of the fundamental processes of grammaticalization, which will serve as a theoretical background for grammaticalization paths that will then be exemplified. I can here hardly even begin to scratch the surface of the topic of grammaticalization, which has received a vast amount of attention. I stress that the characteristics of grammaticalization outlined here are, by necessity, highly simplified. For very thorough introductions to grammaticalization, see, for example Hopper & Traugott (2003), Lehmann (2002) and Heine et al. (1991) as well as the chapters in Narrog & Heine (2011). For a very thorough, but more theoretical discussion on the factors involved in the development of linguistic complexity, see Dahl (2004). An article-length introduction to grammaticalization is, for example, Heine (2003). For a very useful bibliography on grammaticalization made available online by Christian Lehmann, see http://www.christianlehmann.eu/lehr/grammatikalisierung/ (accessed 19 March 2011). A seminal piece of work on grammaticalization and the origin of grammar is Heine & Kuteva (2007), which this section draws on considerably.

13.1.1 The process of grammaticalization

The process of grammaticalization covers various interconnected mechanisms and stages. A lexical item undergoing grammaticalization goes through a stage of **desemanticization**, in that the sum of the semantic content of the lexical item is successively bleached to only the grammatical content of the item. For instance, *be going to* ceases to signify movement and starts to signify intention. With this loss of semantic

content the pragmatic use of the item changes in that it becomes appropriate in a wider set of contexts; it undergoes **extension** or generalization in contexts of use. For example, once *be going to* has started to signify intention and future tense, it may be used in contexts that do not involve any motion (e.g. *I'm going to read out loud to you*). With this, the original morphological and syntactic categorial properties of the item (for example *be going to* being a verb meaning that a mobile and volitional subject engages in movement from place A to place B) tend to erode so that the item is no longer possible to identify as belonging to any major grammatical category (that is, a noun or a verb): the item gets **decategorialized**. This tends to lead to increased frequency in use, which in turn tends to lead to phonological **erosion** (compare *be going to* with *gonna*), which may carry with it certain other morphosyntactic effects, such as loss of status as an independent word (i.e. the item may become a clitic or an affix).[203] The process may thus be said to work along a pathway, or cline, of stages, where the item to the left is less grammatical than its neighbour to the right:

content item > grammatical (functional) item > clitic > inflectional affix[204]

Figure 13.1 The "cline of grammaticality" (Hopper & Traugott 2003: 7 with minimal changes).

It is very important here to keep in mind that these processes and stages are not meant as abrupt and definite. The mechanisms involved are all interrelated, so that it is not possible to draw absolute boundaries between them. A linguistic form changes gradually (even if the speed may vary); it is more like a fluid process with the stages forming landmarks around which features tend to cluster. It is also important to keep in mind that the original source use of the item does not necessarily disappear. It would be quite inaccurate to think of the grammaticalization process as if a linguistic item

203. There is no real way of establishing what comes first, semantic "bleaching" then frequency leading to erosion, or frequency in use leading to "bleaching" and erosion. For instance, the string of words forming the polite inquiry 'How are you?' became a general greeting to the effect that it was no longer a question that needed a meaningful answer, a process sometimes also called **emancipation**. The whole greeting got reduced to 'Hi', which is not answered with 'Fine' anymore in Standard English, although there are varieties where such an answer is appropriate (cf. Bybee 2003b: 155). The question here is which came first: the frequency in use which led to an erosion of meaning and form, or a "bleaching" in meaning which led to a high frequency in use which in turn led to erosion in form. For more on the role of frequency in grammaticalization, see Bybee (2003a). For a collection of articles discussing the interplay of frequency and grammaticalization, see Bybee (2007).

204. It could also be argued that this cline illustrates the path from a more analytic form to a more synthetic form, the further right on the cline the item is.

hops like a frog from one stepping stone to the other. I stress again that this is a fluid process, more a champagne glass tower where the liquid flows in one direction and pools in the glasses along the way. For instance, even though *be going to* has acquired a phonologically reduced form that exclusively functions as a future marker (*gonna*), the source verb is still in use. Due to this linear but overlapping nature of grammaticalization the stages have variously been described as stages along a **pathway**, a **cline**, or a **chain** (where each link of the chain interlaces in the other), among other terms.

13.1.1.1 *Desemanticization*

The term **desemanticization** or "**bleaching**" (sometimes also called **semantic reduction**) tries to capture the fact that when a linguistic construction gets used in new contexts, it has, by then, lost some of its original semantic content. Consider, for example, the construction *be going to* in (358) below:

(358) a. We <u>are going to</u> London.
[MOVEMENT]
S = [+VOLITIONAL; +MOBILE]

b. We <u>are going to</u> have a party.
[INTENTION]
S = [+VOLITIONAL]

c. The trees <u>are going to</u> crack in this storm.
[FUTURE]
S = [−VOLITIONAL; −MOBILE]

In (358a) we have a verb that signifies motion from one place in space to another. It also implies that the subject of the verb would typically be mobile and that the movement would be volitional, something along the line of "We are now choosing to move ourselves in the direction of London". In other words, the construction in (358a) implies an active, agentive verb with an animate subject. In (358b) the inherent semantic feature [+MOVEMENT] of the verb has been lost (there is no movement in space implied), as has the implication that the subject of the verb must be mobile; the construction has been semantically "bleached". In (358c) the construction neither implies movement in space nor intention, and the subject can neither be considered as mobile nor as acting out of volition; the construction may now be used with any type of verb and any type of subject – the construction has been further "bleached". However, as Hopper & Traugott (2003: 94f) point out, what the construction goes through might be better viewed as a shift or redistribution in meaning, rather than loss. We could also analyse it as follows:

(359) a. We <u>are going to</u> London.
[MOVEMENT [INTENTION [FUTURE]]]

S = [+VOLITIONAL; +MOBILE]

b. We are going to have a party.
[INTENTION [FUTURE]]
S = [+VOLITIONAL]

c. The trees are going to crack in the storm.
[FUTURE]
S = [−VOLITIONAL; −MOBILE]

If we think of the construction *be going to* in (359a) as carrying the semantic content of [+MOVEMENT] which in itself carries a notion of [+INTENTION] (the intention of getting to the place the movement will take us to), which in turn has a slight implication of [+FUTURITY] (we haven't yet arrived at the place our movement will take us to), then the "bleaching" process acquires more of a sense of redistribution. In (359a) the most salient semantic feature is [+MOVEMENT]. In (359b) the semantic saliency has shifted to [+INTENTION] while in (359c) it has shifted further, to [+FUTURE]. We may think of is as a shift of perspective where "one meaning is demoted, another promoted" (Hopper & Traugott 2003: 94). Due to this shift in semantic content, the construction can be used in other domains than it could before. Thus it can now be used in contexts that do not involve motion and with subjects that are neither mobile nor carry the semantic feature of [+VOLITION]. This, in a sense, is then a type of pragmatic enrichment, in that the construction gains range of use.

This kind of semantic shift is made possible through something called **pragmatic inference** (Hopper & Traugott 2003), whereby the speaker and hearer communication presupposes inherent inferring. We must always remember that language is used for communication, and that it is shaped according to the needs of that communication. In the case of *be going to*, for example, we might picture that a hearer infers that the external act of movement has an internal motivation of intention and further that the internal motivation of intention implies temporal sequence.

Desemanticization or "bleaching" is not limited to lexical items loosing semantic content, but can also be found in grammatical items losing part (or all) of their grammatical functions. An oft-cited example is that of the nominal endings in Swedish: in Old Swedish the endings were cumulative (portmanteau), expressing case (nominative/accusative/dative/genitive), gender (masculine/feminine/neuter), and number (singular/plural). In Modern Swedish the gender (common/neutral) and number (singular/plural) distinctions are retained, but case has been lost. Notice also that the gender distinctions have shifted from three to two. This is what Norde (2001: 243) terms **functional reduction**. Essentially we are dealing with a type of "bleaching" or shift, similar to the process of semantic reduction described for *be going to* above.

13.1.1.2 *Extension*

Hand in hand with "bleaching" or semantic shift we have **extension**, whereby the "bleached" item gets used in new contexts, the pragmatic enrichment mentioned above. In essence the "bleaching" allows for a context generalization, in that the item can now be used in a more general range of environments. For example, once *be going to* has lost its semantic content of spatial movement, it can be used in a wider set of contexts, such as in (359b) and eventually also (359c). This also has the sociolinguistic implication that an **innovation** is adopted and spreads (**diffuses** or is **propagated**) through the speaking community (Croft 2000). The two parameters of innovation and propagation are equally necessary: an innovation that does not get propagated (or does not diffuse) will not end up entering the linguistic system, but if there are no innovations there is also nothing to propagate.

With the increase of contexts there also tends to be an increase of frequency in use. This in turn may cause further bleaching, which gives rise to even further extension:

> Repetition itself diminishes the force of a word, phrase or construction. Examples are legion. (…) *You guys* generalizes to include females and the word *guy* now can be used in colloquial speech even for inanimate objects. In grammaticalization, the generalization or bleaching of the meaning of a construction is caused by frequency, but it also contributes to additional frequency, as a generalized construction can be used in more contexts, leading to an almost inexorable movement along a grammaticalization path.
> (Bybee 2003b: 157)

For instance the dual effect of "bleaching" and expansion can be seen in the French negative. Originally *ne … pas* could only be used with motion verbs, as in *ne* [*aller*] *pas* 'not [go] a step' but with "bleaching" the construction was extended to usage with any kind of verb as in *ne* [VERB] *pas* 'not VERB'. Further extension and especially an even higher rate of frequency in use has led to the common construction *Ø* [VERB] *pas* 'not VERB', where the original negator has been lost completely and the original noun *pas* 'step' has acquired the meaning of negation. In other words, through the process of "bleaching" and extension, the noun *pas* 'step' has in this context lost the morphosyntactic properties that would categorize it as a noun and has become a functional item (negator) only.

13.1.1.3 *Decategorialization*

Decategorialization is, as mentioned, a consequence of "bleaching" and extension. With the increased frequency in use and the generalization of appropriate contexts, the construction tends to lose the morphosyntactic properties that identify them with a specific category. For example, the noun in the French negative construction *ne … pas* has lost its properties as a noun (something like **ne parler un pas*, with

an article before *pas*, to express 'don't talk' would not be possible). Another typical example of a noun that has decategorialized is the conjunction *while*, which derives from the Old English noun *hwīl* 'length of time'. The word can still be used as a noun, as in *I stayed at home for a while*, but even so some of the typical properties of nouns have been lost. For instance, *while* cannot be inflected for plural, something like **I stayed there several whiles* is not generally acceptable. As a conjunction *while* (as in *while we were eating*) has lost all properties of a noun, as shown by Hopper & Traugott (2003: 107): (*i*) it cannot take articles or quantifiers (**the while we were eating*/**some whiles we were eating*); (*ii*) it cannot be modified by adjectives or demonstratives (**the short while we were eating*/**that while we were eating*); (*iii*) it cannot serve as any other argument of a verb than an object (**while took long*); (*iv*) it can only appear clause initially (compare *The phone rang while we were eating* but not **The phone rang we were eating while*); and (*v*) it cannot be referred to by an anaphoric pronoun. Decategorialization might have a negative ring of loss to it, but in fact the shift also entails gains, in that the item gains the properties of the new category. The functional shift has here led to *while* as a conjunction being used in a way that a noun cannot be used. Consequently, there has been decategorialization leading *while* away from the category Noun, but there has also been a functional gain leading it towards the category Conjunction.

An example of a decategorialized verb is the English modal auxiliary *can*, which derives from the Old English *cunnan* 'to know' and has lost its properties as a lexical verb. For instance, it is no longer possible to use it as a clausal predicate, as in **I can the place* (but cf. the acceptable *I know the place*), or to place it in an infinitival construction **to can* (but cf. *to know*) or in the progressive **canning* (but cf. *knowing*).

Functional, or closed class, items may also undergo decategorialization. For example, demonstrative pronouns form a common source (or origin) for relative clause markers. This is also the case in English, where *that* derives from the demonstrative *that*. As a demonstrative it may inflect for plural (*that book*/*those books*) but as a relative clause marker it has decategorialized – it no longer belongs to the category 'demonstrative pronoun' – and has thus lost some of the properties of that category. It can, for example, not be inflected for plural (cf. *The book that I read* but **The books those I read*).

13.1.1.4 *Erosion*

A linguistic construction that is used very frequently tends to lose phonetic substance, and grammaticalization processes often involve phonetic **erosion** or reduction. We have already seen that the construction *going to* has, as a future marker, eroded to *gonna*. It might in the future further erode to, for example, *gon*. Phonetic erosion can be of various types. Phonetic segments may be lost, even entire syllables. This is what

happened with *going to* > *gonna* (one syllable was lost) and the Old English construction *þā hwīle þe* 'during the time that' which led to the Modern English conjunction *while* (two free words of one syllable each have been lost). Suprasegmental properties may be lost (e.g. stress or tone). This is what happened to the adjective *full* when it grammaticalized to the derivational suffix *-ful* and in the process lost its ability to be stressed (here an analytic construction has become synthetic through grammaticalization). Phonetic autonomy can be lost altogether, whereby the item becomes a clitic or an affix, and may even entail assimilation to the new host. An example of a noun that grammaticalized into an affix is the Proto-Germanic (Indo-European (Germanic)) noun**līka-* 'body, appearance' which came to be used in compound constructions of the type "having the appearance of" (e.g. Germanic **frijōndlika* 'having the appearance of a friend'). The noun eventually "bleached" and extended to become an adjective derivational suffix. In the process it lost in phonological substance, going from Germanic **līka* to Old English *-lic* and, via Middle English *-lich(e)*/*-lik(e)*/*-li(e)* "bleached" and extended further into the Modern English adjective and adverb derivational suffix *-ly*. The adverb suffix might now be eroding even further into complete loss, with colloquial varieties allowing such constructions as *Sit down quiet-Ø* or *It came direct-Ø from him to me* (for *Sit down quiet-ly* and *It came direct-ly from him to me* respectively). Phonetic erosion may also involve phonetic simplification, as exemplified in *going to* > *gonna*, where the cluster /ŋt/ in /gɔɪŋtu/ has been simplified to /n/ in /gɔnə/.

13.1.2 Paths of grammaticalization

The evolution of grammatical categories shows a great deal of cross-linguistic uniformity. That is, the routes by which various grammatical categories enter the linguistic system of a language tend to follow similar patterns among the languages of the world. The following will show some examples of typical grammaticalization paths. For an impressive inventory of grammaticalization paths, based on data from over 500 languages, see Heine & Kuteva (2002), which I base many of the examples in this section on. For the sake of iconicity, I will, essentially, follow the grammaticalization cline in that I will start with the least grammatical linguistic items and end with the most grammatical linguistic items (cf. Figure 13.1 above) that tend to be found cross-linguistically. The examples given in this section are by necessity sketchy and the reader is encouraged to refer to the sources given for further details. A notational remark: '<' means 'derives from' and '>' means 'becomes'.

13.1.2.1 *Content words*

Content items constitute the least grammatical items on the cline. We have seen that nouns and verbs are two near-universal categories (cf. 6.2). Much of the grammatical

material in a language can be hypothesized to ultimately have belonged to one of these two categories. In those languages where adjectives and adverbs constitute their own classes, they are typically ultimately derived from either nouns or verbs.

i. *Adjectives*

Noun > Adjective. Nouns that serve as sources for adjectives tend to be, for example, plants, animals and metals. The English adjective *orange* is an example of this. Nouns denoting gender-specific humans, such as 'woman' and 'man' or 'mother' and 'father', also typically serve as sources for adjectives, most commonly to derive adjectives meaning 'female' and 'male' (compare the English derivation *girly* < $girl_{N}$-*ly*). In !Xóõ, for example, *ạ̀a* 'father' has grammaticalized to mean 'male':

!Xóõ (Khoisan (Southern Khoisan): Botswana)

(360) gùmi ạ̀a
cattle father
'ox' (Traill 1994: 154)

Verb > Adjective. Impressionistically, adjectives seem even more often to derive from verbs. If this is so, one possible explanation might be the fact that cross-linguistically adjectives often follow the *be-X* construction. An example of a very basic descriptive word of size is Swedish *stor* 'big' which ultimately derives from Proto-Indo-European **st(h)ā* 'stand' (Hellquist 1980: 1083). The basic descriptive word of colour, *white*, ultimately derives from the Proto-Indo-European nominal/verbal root **k̂uei-* 'shine; light' (Nielsen 1989: 193).

WIND AND SUN COMPASS

Very common sources for cardinal directions are atmospheric phenomena, especially wind, for 'north' and 'south', and celestial bodies and events, especially the movement of the sun, for 'east' and 'west'. For a very detailed study on the origin of cardinal direction terms, see Brown (1983).

ii. *Adverbs*

Noun > Adverb. Sources for adverbs are often nouns relating to spatial concepts or body parts. For example, the adverb *up* often derives from nouns meaning 'sky' or upper body parts, such as the Teso (Nilo-Saharan (Nilotic): Uganda) *a-kuju* 'Feminine-sky' > *kuju* 'above, up' (Kitching 1915: 74) and the Kupto (Afro-Asiatic (West Chadic): Nigeria) *kúu* 'head' > 'above, up' (Heine & Kuteva 2002: 170 citing Leger 1991: 20).

Notice that the Teso gender marker *a-* 'FEMININE' disappears when the noun has decategorialized to an adverb.

Verb > Adverb. Impressionistically, adverbs more often derive from nouns. However, there are a number of instances where verbs serve as the source of adverbs. It is common that serial verb constructions serve as the source for adverbs, whereby one of the serialized verbs grammaticalizes into a modifier and loses its characteristics as a verb. In Ewe (Niger-Congo (Kwa): Ghana), for example, the verb *dì* 'descend' > the adverb 'down; beforehand' (Lord 1993: 228), probably via the serial *da dì* 'put descend' (cf. Heine & Kuteva 2002: 117).

13.1.2.2 *Function words*

Function items constitute the second level or stage on the grammaticalization cline (content item > function item). They are most typically derived from content items.

i. *Pronouns*

Noun > Pronoun. A very common path of grammaticalization is for the noun 'person' or 'man' to grammaticalize to the impersonal pronoun, for example Latin *homo* 'person, man' > French *on* 'one' (as in 'one should rest after eating'). The noun 'person' often also grammaticalizes into the first person plural pronoun 'we', as in Kono (Niger-Congo (Western Mande): Sierra Leone) where *mɔ̀ɔ̀* 'man, person' > *mɔ̂* 'we. INCLUSIVE' (notice the phonological erosion from long to short vowel and the alternation in tone pattern) (Heine & Kuteva 2002: 233 citing Donald A. Lessau, p.c.). A similar path of grammaticalization is currently taking place in Brazilian Portuguese where the NP *a gente* 'the people' > *a 'ente* 'we' (again, notice the phonological erosion where the initial *g-* is lost; Zilles 2005).

The nouns 'body' and 'head' are common sources for reflexive pronouns (see Schladt 2000 for a survey of sources for reflexives). In Ùsàk Èdèt (Niger-Congo (Cross River): Nigeria) *únem* 'body' > reflexive marker 'self' (Essien 1982: 98).

Verb > Pronoun. This concerns mainly demonstratives, which often derive from general verbs of motion meaning 'come' or 'go', as in Mopun (Afro-Asiatic (West Chadic): Nigeria), where *sò* 'this' < Proto-Chadic (Afro-Asiatic (Chadic)) **(-)sò* 'to come' and *dì* 'there' < Proto-Chadic **dò* 'to go' (Frajzyngier 1987). However, Diessel (1999a) argues that demonstratives actually represent a category of function words that has not grammaticalized from any lexical source but that they in fact belong to very basic original language. He points out the lack of consistent evidence for lexical sources for demonstratives, as well as the rather systematic sound symbolism that demonstratives display (especially with regard to vowel quality: proximate demonstratives tend to have a higher pitch than distal demonstratives, as in *this* (proximate), which has a high vowel, versus *that* (distal), which has a low vowel), and given that

"sound symbolism is usually associated with newly created words" (Diessel 1999a: 151f) this might serve as an indication that these are, in fact, typically original, non-grammaticalized words.

ii. *Adpositions*

Noun (spatial) > Adposition. Spatial nouns are very common sources for adpositions. The Lezgian noun *pad* 'side' > the postposition *patal* 'for' (Haspelmath 1993: 222). The Swahili noun *m-paka* 'CLASS.3-border' > the preposition *mpaka* 'until' (Heine & Kuteva 2002: 61). Body parts also often serve as sources for locational adpositions. In Wan (Niger-Congo (Eastern Mande): Ivory Coast), for example, the noun *káò* 'belly' > the postposition 'inside' and the noun *dìŋ* 'hip, side' > 'near, close.to, next. to' (Nikitina 2008: 35).

Verb (motion) > Adposition. This seems slightly less common than adpositions deriving from nouns. Typically, general movement verbs may grammaticalize into adpositions, often via verb serialization. Examples are the Ewe verb *ná* 'give' > 'for' (Heine & Reh 1984: 37). The North-Central Malaitan **fa'asi* 'leave, forsake' > Toqabaqita *fa'asi* 'from' (Lichtenberk 1991), or the French *arriver* 'arrive' > Haitian Creole *rivé* 'to' (Heine & Kuteva 2007: 73).

iii. *Articles*

Demonstrative > Definite article. Definite articles "are almost invariably derived from demonstratives" (Heine 2003: 594), as is the case in English with English *the* < *that*. In Basque the oblique stem *har-* of the distal demonstrative *(h)ura* 'that' > *-a*/*-ak* 'the' (singular/plural; Hualde & Ortiz de Urbina 2003: 119, 123); again, notice the phonological erosion of *har-* to *-a* (*-k* is a plural suffix) and that we have an instance of an analytic construction (the free word demonstrative *(h)ura*) developing into a synthetic one (the definite suffix). For a lengthy discussion on demonstratives and how they tend to be grammaticalized, see Diessel (1999b).

Numeral 'one' > Indefinite article. Indefinite pronouns almost invariably derive from the numeral 'one', as is the case in English: *a(n)* < Old English *ān* 'one'. In Turkish *bir* 'one' is used both as a numeral and as an indefinite article: *bir elma* 'one/an apple' (Kornfilt 2003: 106).

THE FUTURE SUN

In Mandinka (Niger-Congo (Western Mande): Senegal) *sina* means both 'sun' and 'future tense' via a grammaticalization path of *si-na* 'sun-come' > 'tomorrow.' (Claudi 1994)

iv. *Auxiliaries*

Verb > Auxiliary. This is an almost universal path of grammaticalization. TMA auxiliaries almost invariably derive from verbs. For instance, future tense auxiliaries very often come from verbs meaning 'want' or general motion verbs such as 'go'. For example, the English *will* 'FUTure' < Old English *willan* 'to want' and the Klao *mu* 'FUTure' < 'go':

Klao (Niger-Congo (Kru): Liberia)

(361) a. ɔ̄ɔ̄ **mū** nī tó
he.IPFV go LOC store
'He is going to the store.'

b. ɔ̄ɔ̄ **mū** nī kpa
he.IMP FUT water hit
'He will swim.' (Marchese 1986: 74)

In (361a) the verb *mū* 'go' functions as a lexical verb, while in (361b) it has been decategorialized to a future marking auxiliary. Progressive and imperfective aspects very often derive from stative or copular location expressions (*X stands at Y* or *X is at Y*). Progressives are then in turn very often the source for imperfective aspect and present tense markers. In Maninka, for example, we find the following grammaticalization path: *yé* 'to see' > copula > imperfective via a kind of exclamatory construction of the type "See X in the bush!" (Kastenholz 2003: 39):

Maninka (Niger-Congo (Western Mande): Guinea)

(362) a. kamori **yé** wàá dɔ́
PN COP (< 'see') bush in
'See Kamori in the bush!' > 'Kamori is in the bush.'

b. ń **yé** màló dɔ́mun-na
1SG IPFV rice eat-IPFV
'I am eating rice.' / 'I (usually) eat rice.' (Kastenholz 2003: 40)

In (362a) the verb *yé* 'see' already serves as a copula, while in (362b) it has become an imperfective marker.

Verbs meaning 'finish' are very often the source for completive, perfective or perfect markers Bybee et al. (1994). Modality markers often derive from verbs meaning 'to be able to' or 'to know' to form dynamic event modalities (such as the English *can* < Old English *cunnan* 'to know'). For a classic piece of work on the grammaticalization paths of TMA categories, see Bybee et al. (1994).

Negative auxiliaries often derive from verbs meaning 'leave' or 'lack', such as the Klao *sē* 'NEGATOR' < *sɨ* 'let go, leave' (Marchese 1986: 173, 182). Negative existential verbs ('to not.exist') form another very common source for negators (cf. Croft 1991);

these may in turn actually derive from a negator plus an existential or locative, which in turn may have derived from verbs for 'live', 'sit', 'stand', and the like. One may hypothesize a path of the following kind: verb ('live'/'sit'/etc.) > existential/locative > negator+existential/locative > negative existential verb > negative auxiliary.

v. *Conjunctions*

Noun > Complementizer. It is very common for general nouns such as 'thing', 'matter', 'place' and so on to serve as sources for complementizers. For example, the Japanese *koto* 'NOMINALIZER' < 'thing' (Lehmann 2002: 56).

Verb > Complementizer. Verbs for 'say' or 'be like'/'resemble' form another very common source for complementizers, such as the Taiwanese (Sino-Tibetan (Chinese): Taiwan) verb *kóng* 'say' > complementizer ((363) below) and the Twi (Niger-Congo (Kwa): Ghana) verb *sɛ* 'resemble, be.like, etc.' > complementizer ((364) below):

Taiwanese (Sino-Tibetan (Chinese): Taiwan)

(363) a. a i **kóng** lì àn-noâⁿ?
PRT 3SG say 2SG how
'So what did he say about you?'

b. i kanna liām **kóng** á-hô thàn chiok-chē chîⁿ
3SG just nag that PN earn much money
'He kept insisting that A-Ho earns a lot of money.' (Chappell 2008: 66, 71)

Twi (Niger-Congo (Kwa): Ghana)

(364) a. kofi **sɛ** amma
PN be.like PN
'Kofi resembles Amma.'

b. na ama nim **sɛ** kofi bɛyɛ adwuma no
PAST PN know that PN did work ART
'Ama knew that Kofi had done the work.' (Lord 1993: 152, 159)

It seems near at hand to hypothesize that the grammaticalization path went along the lines of verb ('say'/'resemble') > quotative > complementizer. In fact the English *like* seems to be grammaticalizing into a quotative marker, especially in American English, as in *He said, 'What's the matter with you?' And I'm* ***like****, 'None of your business'* (cf. Romaine & Lange 1991). One day it might end up grammaticalizing into a complementizer.

Demonstrative > Complementizer. This is also very common and is what happened in English with the complementizer *that* < the Old English demonstrative *þæt*. Complementizers often go on to grammaticalize into relative clause markers.

vi. *Classifiers*

Noun > Classifier. Classifiers frequently have their origin in nouns, typically generic nouns (such as 'human', 'tree', 'animal', 'seed' and so on), which commonly serve as sources for sortal classifiers. In Mandarin, for example, the noun *tóu* 'head' > classifier for big animals such as oxen, pigs, elephants (but also, interestingly, garlic) and *zhī*[2] 'twig' > classifier for long and slim (essentially one-dimensional) tools, such as arrows, chopsticks, flutes, rifles and so on (Bisang 1999: 133). Body parts may serve as sources for mensural classifiers.

13.1.2.3 *Clitics*

Clitics constitute the third level or stage on the grammaticalization cline (content item > function item > clitic). From this stage onwards the concern is mostly about how the grammaticalization affects the morphosyntactic and phonological form of the item. Particles, adpositions, pronouns, articles and auxiliaries are examples of items that may get reduced in form to the extent that they need a host to cliticize to. Recall that clitics are syntactically independent, even though they are phonologically dependent items. The erosion has thus not affected the syntactic dependency, only the phonological one.

For example, the late Imperial Latin construction involving the inflected auxiliary 'have' developed into the Old Spanish future marker clitic:

Latin (Indo-European (Italic): present-day Italy)

(365) vidēre **habēo**
see.INF have.1SG
'I have to see.' (Hock 2003: 450)

Old Spanish (Indo-European (Romance): present-day Spain)

(366) a. veer=**he**
see=FUT.1SG
'I will see.'
b. veer=lo=**he**
see=it=FUT.1SG
'I will see it.' (Hock 2003: 450)

In (365) we have the Latin construction with an inflected auxiliary following a lexical verb. In (366) the auxiliary has eroded to a clitic; it needs a host to attach to. Notice that it is still syntactically independent from the host and may attach to any element, as shown in (366b) (in this case another clitic). The clitic further developed into an inflectional affix, see below.

Another oft-cited example of cliticization is the Proto-Norse (Indo-European (Germanic)) reflexive pronoun **sik* 'self' that eroded to a clitic *-sk*, which is still

traceable in, for instance, the English verb *bask* < Old Norse *baða-sk* < Proto-Norse **baða sik* 'bathe (one)self'. This clitic also further developed into an affix, see below.

13.1.2.4 *Inflectional affixes*

Affixes constitute the fourth level or stage on the grammaticalization cline (content item > function item > clitic > affix). At this stage the item has become both phonologically and syntactically dependent on its host. For example, the Old Spanish clitic mentioned in (366) above developed into an inflectional future affix in Modern Spanish:

Modern Spanish (Indo-European (Romance): Spain)

(367) a. ver-**é**
see-FUT.1SG
'I will see.'

b. lo ver-**é**
it see-FUT.1SG
'I see it.' (Hock 2003: 450 with minimal changes)

Notice that the affix is now syntactically dependent on its host. It could not, for example, attach to the object of the clause; something like **ver-lo-é* or **lo-é ver* would not be possible for 'I see it'. The Proto-Norse **-sik* > *-sk* 'self' grammaticalized further to a derivational affix *-st*[205] which in turn grammaticalized to the passive affix *-s* (cf. Enger 2002), as in Swedish *tala* 'speak' (active) versus *tala-**s*** 'be spoken' (passive). These are again examples of analytic constructions becoming synthetic (cf. the analytic *vidēre habēo* > the synthetic *ver-é*).

Case markers often ultimately derive from nouns or verbs. General motion verbs may develop into, for example, ablative, allative and benefactive cases, while general verbs for 'give' often develop into benefactive and ultimately dative cases. For example, the Sa'a (Austronesian (Oceanic): Solomon Islands) dative case marker *huni-* < Proto-Oceanic (Austronesian (Oceanic)) **pa(nñ)i* 'give' (Lichtenberk 1985: 25). General locational or spatial nouns such as 'back', 'front', 'side', etc., may develop into locational case markers, often via an adpositional stage.

13.1.3 A very short note on unidirectionality and degrammaticalization

Grammaticalization is overwhelmingly **unidirectional** in that we have many known cases of items travelling from the left to the right on the grammaticalization cline, i.e. going from less grammatical to more grammatical, but only very few examples

---

205. Thus stative verbs could be derived with *-st* from nouns and adjectives, such as *flokkr* 'crowd' (noun) > *flykkjast* 'flock (together)' (verb) and *grænn* 'green' (adjective) > *grænast* 'become green' (Enger 2002: 84).

of the reverse, i.e. items going from more grammatical to less grammatical. The term **degrammaticalization** was coined by Christian Lehmann (1982) to capture what he postulated as the impossible concept of reversed grammaticalization. For instance, while we rather often see that a lexical item such as the Proto-Germanic noun**līka-* 'body, appearance' grammaticalizes into a derivational suffix such as the English *-ly*, we rarely see instances of where the reverse happens. In other words, it is not generally expected that, for example, the English derivational suffix *-ly* will degrammaticalize and become a content word. Since Lehmann's coinage there has been a lively discussion whether or not the assumption of unidirectionality in grammaticalization holds. It is beyond the scope of this chapter to give a thorough discussion of the validity or invalidity of unidirectionality and the putative counterexamples to it. For accessible overviews of the arguments against unidirectionality, see Haspelmath (2004) and Janda (2001) with references.

Some commonly cited examples of degrammaticalization (or 'antigrammaticalization') are the following (for more examples of attested and valid cases of degrammaticalization, see Haspelmath 2004 and Traugott 2001):

i. The genitive suffix *-s* in English and Scandinavian languages degrammaticalizing to a clitic *=s* (but see Traugott 2001 for an argument against this being a legitimate example of degrammaticalization). (Norde 2001)
ii. The Pennsylvania German (Indo-European (Germanic): USA) auxiliary in the preterit subjunctive form *welle* 'would' developed into a lexical verb *wotte* 'wish, desire'. (Burridge 1998)
iii. The Irish first person plural subject suffix *-mid*/*-muid* developed into an independent first person plural pronoun *muid(e).* (Bybee et al. 1994)

13.2 Contact-induced change and linguistic areas

Languages may also change due to contact with other languages. Language contact in essence means that a number of speakers of one language have a certain degree of competence in another language (or several other languages). I stress once again that when we deal with languages we must always remember that they are dynamic systems used by speakers as a tool for interaction. It would be a very rare community indeed that was self-sufficient to the extent that it never had contact with any other community at all. We may therefore assume that most, if not all, communities (and hence the languages of these communities) have been in contact with other communities (and their languages). It is, in fact, rather more common in the world that speakers live in multilingual settings, where they use different languages for different purposes or in different situations, than that they live in monolingual settings where

one and the same language is used for all purposes and in all situations. However, the extent to which this contact will influence any or all of the languages in question may differ considerably.

This section will briefly outline some of the sociolinguistic factors and mechanisms involved in contact-induced change. It will also mention what kinds of linguistic features languages may adopt from other languages as well as introduce and briefly discuss the concept of linguistic areas. As indicated by the recurrent word 'briefly', this section can by no means cover all the various issues related to language contact, linguistic areas, and contact languages. Each of these is a vast subject which can only be very shortly summarized here. A very accessible introduction to contact linguistics is Thomason (2001). For article-length overviews of contact-induced change, see Koptjevskaja-Tamm (2010) and Aikhenvald (2006a). The latter is a comprehensive introduction to Aikhenvald & Dixon (2006a), which contains a number of chapters discussing various contact phenomena in different languages and areas. A very detailed overview of grammatical borrowings can be found in Matras & Sakel (2007), which functions as a survey of contact-induced change in different languages and is based on Jeanette Sakel and Yaron Matras' *Language Convergence and Grammatical Borrowing Database* (henceforth LCGB). A detailed survey of the typology of loanwords can be found in Haspelmath & Tadmor (2009), which is based on the *World Loanword Database*. A classic in contact linguistics is Thomason & Kaufman (1988).

A terminological note: the language from which a linguistic feature originates is here called the **source language**, while the language which adopts a linguistic feature from another language is here called the **recipient language**.

13.2.1 Sociolinguistic factors and mechanisms of contact-induced change

The type of contact situation, as well as such things as the size and language attitudes of the communities in question, all combine to lead to different kinds of borrowing, or **diffusion**. Linguistic features can be taken over wholesale, where both the form and the function are borrowed, or the principles can be taken over, where the function, but not the form is borrowed. Or anything in between.

13.2.1.1 *Types of contact situation*

Very simplified, the sociolinguistic factors involved in contact-induced change boil down to the basic parameter of dominance. Essentially, there are two types of contact situations with respect to dominance: (*i*) the language(s) in contact are on a more or less equal footing, and (*ii*) the language(s) in contact are not on an equal footing, i.e. the contact situation involves one (or occasionally several) language(s) that dominate over other languages. Dominance can be due to factors such as social prestige

or political power. The two basic situations are sometimes termed **balanced** versus **displacive** language contact (cf. Aikhenvald 2006a). In both cases we tend to get bi- or multilingualism. However, in a balanced language contact situation the multilingualism tends to be mutual in the sense that speakers and languages influence each other and speakers of both or all the languages in contact may be bi- or multilingual. In other words, balanced contact situations tend to lead to a two- (or several-) way influence. In the Vaupés river basin area of Amazonia at the Colombian/Brazilian border, for example, speakers of the same language are considered blood relatives. Hence spouses are sought from other language groups. This linguistic exogamy has given rise to a situation of stable multilingualism, where borrowing is more or less multidirectional (Aikhenvald 2006b).

In a displacive language contact situation the influence tends to be one-way: the multilingualism is typically found only among the speakers of the dominated language(s) and not among the speakers of the dominating language. There are, for example, no monolingual Welsh speakers in Wales; any Welsh speaker will be at least bilingual in English (although in some rural areas some older speakers may be more fluent in Welsh than in English). There are, however, a vast number of monolingual English speakers that do not need to have any knowledge of Welsh to get by. Thus the two types of situation may lead to radically different results: while in balanced language contact situations the linguistic systems tend to enrich each other – forms from the various languages are incorporated into the other languages, ultimately leading to a more complex system – displacive language contact systems rather lead to the opposite, in that the dominated language(s) may adopt linguistic structures of the dominating language. Instances of extreme dominance may ultimately lead to language loss if speakers end up shifting to the dominating language (**language shift**). This may come about due to enforcement, such as laws against using minority languages, or due to lack of socioeconomic opportunities in the minority language, leading the speakers of the minority language community to opt for the dominating language (cf. 3.1.1).

Note that a language can be both dominating and dominated at the same time: Katanga Swahili (Niger-Congo (Bantoid)), for example, is the lingua franca of the North-Eastern region of The Democratic Republic of Congo. It is taught in primary schools and is recognized as a national language. As such it is a language dominating over the regional languages in the area. However, Katanga Swahili is in turn dominated by French, the administrative language of the country (see further de Rooij 2007).

13.2.1.2 *Types of borrowing*

Linguistic features may transfer between languages in various ways. Strictly speaking diffusion of linguistic features is not a matter of borrowing, but rather one of copying, as Lars Johanson rightly points out (see e.g. 2008): speakers do not "borrow" features which they use for a while and then return to the "owner" or the source language, but copy features from one language (or variety) into another. As the term 'borrowing' is so widely used, I will conform to that convention here. However, the reader should at all times keep in mind that linguistic 'borrowing' is a matter of code copying rather than anything else.

Following Matras & Sakel (2007) I distinguish between **matter borrowing** (or MAT), where the phonological shape and the morphological form is taken over from one language to another, and **pattern replication** (or PAT), also called **calques**, where effectively the function of the linguistic feature is taken over but not the form. This essentially mirrors the distinction between 'fabric' and 'pattern' borrowing as previously outlined in Grant (1999 and 2002). An example of a matter borrowing is the Spanish (source) word *entre* 'between' in Guaraní (Tupian (Tupi-Guaraní): Paraguay) (recipient), where both meaning and form have been taken over in the recipient language (Matras 2007: 42). An example of a pattern replication can be found in the Estonian (Uralic (Finnic): Estonia) (recipient) calques from German (source), as in *läbi-hammustama* 'bite through' (lit. through-bite) < German *durch-beißen* 'bite through' (lit. through-bite), where the pattern has been taken over but not the form (Aikhenvald 2006a: 25). Matter borrowing typically entails at least a partial overlap with pattern replication in that when matter is borrowed at least some of the functions of the linguistic forms are also taken over. A rare example of matter borrowing without any pattern replication is the German word *Handy* 'mobile phone': the word is supposedly a loan from English but, if so, the meaning has changed (cf. Sakel 2007: 26). A similar example is the French noun *footing* 'jogging' (Curnow 2001: 427), which again is in effect an English word but with an entirely different meaning.

Very broadly speaking, matter borrowing tends to be more frequent in the recipient language the more dominant the source language is. In other words, the more displacive the contact situation, the more matter borrowing. Conversely, the more balanced the contact situation, the more pattern replication we tend to get. This is of course a gross simplification. Other highly relevant parameters are, for example, language attitude. If a community is averse to incorporating 'foreign' elements in their language, pattern replication will be more prominent than matter borrowing. There was, for example, in the early twentieth century, a conscious effort to make Estonian less German-like, which resulted in considerable restructuring in both lexicon and grammar (Aikhenvald 2006a). Here the matter borrowings were more straightforward to weed out, while the less obvious pattern replications were more elusive to these efforts.

These kinds of attempts at purism often, but by no means always, emerge when a community feels under threat; the language becomes a vital sign of identity, and speaking a 'pure' form of the language becomes an important statement of loyalty.[206]

13.2.1.3 *Mechanisms of contact-induced change*

The mechanisms by which new material enters a language may differ. Typically, but not always, it starts with **code-switching**. Code-switching essentially means that speakers use more than one language or variety in the same utterance, a rather common phenomenon with bi- and multilinguals. Some examples of code-switching made by myself between my own two mother tongues are (368a–c), where the code-switched material is in italics.

Swedish/English code-switching

(368) a. page forty-*nio* to *sjuttio*-eight
-nine seventy-
'Page forty-nine to seventy-eight.'

b. det *make*-ar ingen *sense*
it make-PRES INDEF.PRON sense
'It doesn't make any sense.'

c. jag vet inte om jag hinner men *I'll try*
1SG know NEG if 1SG make.it but
'I don't know if I'll make it, but I'll try.' (source: personal experience)

In (368) my utterances contain both Swedish and English. The code-switched material, the material in a different language from the one I started the utterance in, is in italics. Notice that in (368b) I ended up inflecting the code-switched verb with a Swedish verbal inflection (*make-ar*, pronounced /meɪkaɹ/).

When a sizeable proportion of the speakers engage in code-switching, especially systematic code-switching, where certain features regularly get code-switched, linguistic features may diffuse from one language to another. In other words, the feature

---

206. An interesting confirmation of how bilingualism promotes matter borrowing over pattern replication are the findings of Brown's (1989) cross-linguistic survey of 148 languages on names for weekdays. He found that in those societies where bilingualism had been promoted, for instance in colonies where the indigenous population had been integrated in colonial culture to a certain extent, weekday names were typically borrowed from the colonizers. This is especially evident in South America, with weekdays typically having been borrowed from Spanish. However, in those colonial settings where the indigenous population had generally not been integrated in colonial culture, and where consequently bilingualism had been much more limited, weekday names had been created by new coinages using only native linguistic materials. This is evident in North America, where weekdays typically involved native coinages. For more details, as well as a hierarchy of weekday borrowings, see Brown (1989).

may propagate. For more on the linguistic effects of code-switching, see, for example Myers-Scotton (2002) with further references. For a very enjoyable exposé on various myths and beliefs about multilingualism in general, see Cruz-Ferreira (2010).

A **transferred** form is essentially borrowed wholesale from one language into another. The matter borrowing of Spanish *entre* into Guaraní mentioned above is an example of a transferred form. Bound morphemes may also be transferred, as the English suffix *-able* < Old French.

An already existing tendency may be reinforced, or **enhanced**, by language contact. This is especially true of displacive contact situations. In other words, if a dominated language already has a tendency towards a construction, and the dominating language has the same or a very similar construction, this construction is likely to become reinforced in the grammar of the dominated language. In Pennsylvania German, for example, already existing tendencies are 'helped along' by intense contact with American English. Thus the common English '*get*-construction' where *get* has an auxiliary function that does not mean 'receive' has reinforced a similar construction with *kriege* 'to get, receive' in Pennsylvania German:

Pennsylvania German (Indo-European (Germanic): USA)

(369) a. mir kriege gesaagt
we get told
'We get told.'

b. hen dir sel geduh geriegt
have you that done got
'Did you get that done?' (Burridge 2006: 186f)

In (369) we have the construction *kriege* + past participle, where *kriege* functions as an auxiliary in the same way as in the English '*get*-constructions' (cf. the translations of the examples). Enhancement is an example of pattern replication, where the pattern is adopted but not the form.

A similar process is **contact-induced grammaticalization**, where a lexical item in the recipient language grammaticalizes to match a grammatical category in the source language (i.e. a pattern replication of a grammaticalization path). In Basque, for example, the verb *eraman* 'carry' has taken the same grammaticalization path as Spanish *llevar* 'carry' and grammaticalized into a perfect progressive marker (expressing 'to have been doing something') (Jendraschek 2006: 157). Here the recipient language (Basque) follows the path of the source language (Spanish). The recipient language may also by way of grammaticalization create a new category to match an existing category in the source language. In the Arawakan language Tariana, for example, the verb *-sita* 'finish' has grammaticalized into a perfective aspect marker matching existing categories in the East Tucanoan languages with which it is in contact (Aikhenvald 2003: 12).

An already existing feature may also get **extended** in its function or use as an effect of language contact. In Mosetén (Mosetenan (Mosetenan): Bolivia), for example, inanimates do not generally get marked for plural. This is, however, changing due to the intense contact with Spanish, where animacy plays no role in the plural marking of nouns (Sakel & Matras 2008). In this case the already existing plural marking of nouns in Mosetén gets extended in use to all nouns. This is again an instance of pattern replication, where the pattern of the dominating source language (Spanish) is being taken over in the dominated recipient language (Mosetén).

An already existing linguistic feature that is similar in form to a linguistic feature in the source language may get reinterpreted to take over the functions of the form in the source language. Thus the recipient language grammatically **accommodates** its feature to the functions of the similar sounding feature in the source language. In Likpe (Niger-Congo (Kwa): Ghana), for example, TMA is marked by prefixes on the verb, except for the present progressive which is expressed analytically. This is due to intense contact with Ewe: the Likpe verb *lɛ́* 'hold' sounds very similar to the Ewe progressive marker *le* 'be.at:PRESent' and has accommodated to become a present progressive marker. Compare the two constructions in Likpe and Ewe:

Likpe (Niger-Congo (Kwa): Ghana)

(370) ɔ-lɛ́ ka-mɔ bo-té
3SG-hold CM-rice CM-sell
'She is selling rice.' (Ameka 2007: 112)

Ewe (Niger-Congo (Kwa): Ghana)

(371) é-le mɔ́lu dzráx-ḿ
3SG-be.at:PRES rice sell-PROG
'She is selling rice.' (Ameka 2007: 112)

Due to its superficial similarity in form to Ewe *le*, the Likpe *lɛ́* 'hold' has been grammatically accommodated to have the same function as Ewe *le* in the source language pattern, even though the semantics differ. This is essentially also a case of pattern replication, although one could possibly argue that the similarities in form between the feature in the recipient and the source languages trigger a passive kind of matter borrowing.

Calques, or **loan translations**, are, as mentioned above, instances where the recipient language copies a pattern of the source language but not the form. The Manange (Sino-Tibetan (Bodic): Nepal) expression ²*ta* ³*pi-nʌ* 'because' (lit. 'what say. EVIDential') is a calque from Nepali *kina bhane* 'because' (lit. 'why say') (Hildebrandt 2007: 293). Here, again, we have a case of pattern replication without any borrowing of form. Borrowings may sometimes occur parallel to the native expressions,

sometimes called **double marking**. An example of double marking is the Mezquital Otomí (Oto-Manguean (Otomian): Mexico) construction for repetition of an event, where the native form [main verb + *ma ʻnagi* 'again'] occurs parallel to the calqued expression [*pengi* 'return' + main verb] < Spanish [*volver a* + main verb] (as in *Volvió a salir* 'He went out again' lit. '(re)turn.3SG.PRETerit to go.out'):

Mezquital Otomí (Oto-Manguean (Otomian): Mexico)

(372) **dá=pengi** dá=uni ma ʻnagi
PAST.1SG=return PAST.1SG=give again
'I gave it to them again.' (Hekking & Bakker 2007: 456)

In (372) the calqued construction (in bold) is used alongside the native construction. Double marking, or parallel constructions, may serve to give 'foreign' constructions an air of legitimacy and thus may ultimately lead to the integration of the construction in the recipient language.

Borrowed matter may be **reinterpreted** and acquire new functions in the recipient language. In Hup, for example, the borrowed matter *ni-* 'be' from Tukano *diî* ([nii]) 'be' is not only used for inferred evidentials, as in Tukano, but also as a verbalizer, a function for which Tukano has a different marker (Epps 2007: 558). Acholi (Nilo-Saharan (Nilotic): Uganda) is an example of a language where borrowed matter has been completely reinterpreted: the Karimojong gender markers *á-* (feminine) and *é-* (masculine) have been reanalysed as the Acholi number markers *à-* (singular) and *è-* (plural). Compare the two languages in the example below:

Karimojong (Nilo-Saharan (Nilotic): Uganda)

(373) a.

SINGULAR	PLURAL		
ɛ́-kìl-é	ŋí-kíl-yók	'man'	(MASCULINE)
á-bɛ́ɛ́rú	ŋá-berò	'woman'	(FEMININE)
í-kɔ̀kʊ	ŋí-dwé	'child'	(NEUTER)

Acholi (Nilo-Saharan (Nilotic): Uganda)

b.

SINGULAR	PLURAL		
à-ṯín	è-ṯîn-ɔ	'child'	
à-ṯín-ḏyang	è-ṯîn-ḏok	'calf'	
à-kwɔ́	è-kúw-é	'thief'	(Storch 2006: 104f)

As can be seen in Example (373), the gender markers of the Karimojong source language have been borrowed in form but have been reinterpreted as number markers in the recipient language Acholi.

13.2.2 Feature 'borrowability'

In principle any linguistic feature can diffuse from one language to another (or several others). However, it seems as if some linguistic features are more prone to diffusion than others. This section will give some examples of features that tend to diffuse. It relies heavily on the findings of LCGB as discussed in Matras (2007).

13.2.2.1 *Phonology*

Typically through introduction of loan words, the phonology of the recipient language may be altered. Most common seems to be the addition of phonemes, such as /m/ in Basque which entered the language first via Celtic loans and was then reinforced by subsequent loans from Latin (Trask 1998).[207] Consonants seem more prone to diffusion than vowels, which is probably due to the simple fact that consonant inventories are typically larger than vowel inventories. Suprasegmentals, especially tone, also commonly diffuse. Thus many of the Tibeto-Burman languages that have been or are in contact with Chinese languages (those in the 'Sinosphere') developed tone, as opposed to those Tibeto-Burman languages in the 'Indosphere', i.e. in the Indic language area (LaPolla 2001: 234f). In fact there seems to be a tentative tendency that suprasegmentals diffuse somewhat easier than segmental phonemes (Matras 2007: 39).

13.2.2.2 *Morphology*

Bound morphemes may be borrowed, as was the case with the Acholi number markers in Example (373). Derivational morphemes seem to diffuse easily. For example, the LCGB shows borrowing of a diminutive suffix into Macedonian Turkish, Yiddish (Indo-European (Germanic): Central Europe), Imbabura Quechua, Purépecha, and Rumungro (Indo-European (Indic): Czech Republic) (Matras 2007: 43).

Contact may also cause languages to move from one morphological type to another. For instance, Indonesian has become increasingly more isolating under the influence of the various lingua franca varieties of Malay (Tadmor 2007: 308), while the largely concatenative Likpe is increasingly relying on reduplication under influence from Ewe (Ameka 2007: 110).

13.2.2.3 *Nominal categories*

Number marking seems to diffuse easily, while for example gender and case markers do not. In fact, there are no instances of borrowed case markers in the LCGB (Matras

---

207. In fact, internal tendencies were also in place, whereby /b/ > /m/ and /nb/ > /m/. Contact thus not only brought new lexicon with the phoneme into the language, but also reinforced an already existing tendency. For more details see Trask (1998: 317f).

2007). However, Northern Tajik (Indo-European (Iranian): Tajikistan), which is not in their sample, borrowed almost all case markers from Uzbek (Altaic (Turkic): Uzbekistan) (Doerfer 1966: 95). Classifiers seem susceptible to language contact; Hup and Cantonese acquired systems similar to the languages they are in contact with (Aikhenvald 2006a: 16), while the originally Bantu noun class system seems to have diffused widely to other Niger-Congo languages and even beyond, to, for instance, the Nilotic (Nilo-Saharan) language Luo in Kenya (Dimmendaal 2001: 377ff). While Luo does not have classifier agreement marking on nominal modifiers, the Bantu markers entered the language through extensive borrowing where the noun-class prefixes were retained. Additionally, a language internal process, possibly reinforced by the Bantu lexical influence, is seeing lexical heads develop into noun-class prefixes (Dimmendaal 2001: 382).

13.2.2.4 *Verbal categories*

When it comes to verbal categories, especially TMA, there seems to be a hierarchy of how categories are affected by language contact:

modality > aspect > future tense > (other tenses) (Matras 2007: 46)

The further left on the hierarchy, the more likely it is that the category is prone to contact-induced change, and the more likely matter borrowing is. In other words, modality is most likely to be affected by language contact while tenses other than the future are the least likely to be affected by language contact. Matter borrowing also decreases in frequency further right on the hierarchy – contact-induced changes in the future tense are thus likely to be pattern replications and not matter borrowing. This typically also works as an implicational hierarchy in that if a language has contact-induced aspect marking, then it will also have contact-induced modality, and so on. There are, for example, numerous cases of diffused evidentials and markers of obligation, such as the tense-evidential system of Tariana enclitics modelled on the East-Tucanoan suffixes (Aikhenvald 2003) and the obligative *tiene que* 'must' in Rapanui from Spanish (Fischer 2007: 391). Similarly, progressives and completives seem to diffuse fairly easily, such as the Likpe progressive construction in (370) above and the Indonesian completive *sudah* < Sanskrit *śuddha* 'cleansed, cleared, acquitted' (Tadmor 2007: 315). For more examples, see Aikhenvald (2006a) and Matras (2007).

Contact-induced change in valency generally involves pattern replication, as in the Purépecha periphrastic passive modelled on the Spanish *ser* ('be')-construction (Chamoreau 2007: 469); typically contact-induced change in valency involves reinforcement and extension of already existing features (Matras 2007).

13.2.2.5 *Parts-of-speech*

Nouns are more readily borrowed than verbs (Matras 2007), although there are numerous examples of verb borrowings.[208] When it comes to functional word classes, adpositions seem to diffuse with relative ease, as, for example, 'between', which is the most frequently borrowed adposition in the LCGB with Indonesian *antara* (< Sanskrit), Guaraní, Purépecha and Mezquital Otomí *entre* (< Spanish), Domari (Indo-European (Indic): Middle East) *bēn* (< Modern Standard Arabic) (Matras 2007: 42).

Numerals may also be affected by contact-induced change. In fact, more than two thirds of the languages in the LCGB have some kind of matter borrowing of numerals (Matras 2007: 50). Jaminjung (Australian (Jaminjungan): Australia), for example, has borrowed numerals higher than '3' from Kriol (Creole (English-lexified): Australia; Schultze-Berndt 2007: 381), while Katanga Swahili uses French numerals for dates (de Rooij 2007: 130).

Pronouns are often thought to be quite resistant to diffusion, although a number of cases of borrowed pronouns are known. The English *they*, *them*, *their*, for example, were borrowed from Old Norse *þeir* (nominative), *þeim* (dative), *þeir(r)a* (genitive) due to intense Scandinavian contact and replaced the Old English *hi/hie*, *him*, *hira/hiera* respectively (see, for example, van Gelderen 2006: 57, 98). For more examples on pronoun borrowing and a discussion on the origin of the Pirahã pronouns in particular, see Thomason & Everett (2005).

Conjunctions are extremely prone to diffusion. All the languages in the LCGB have borrowed conjunctions. There seems to be a strong indication for an implicational hierarchy:

but > or > and (Matras 2007: 54)

While some languages in the database borrow only 'but' and 'or', none of them have borrowed 'and' without having borrowed the other two conjunctions on the hierarchy.

Discourse markers also diffuse extremely easily, and while it is not an absolute universal, it is fair to say that it is one of the categories most likely to be borrowed

---

208. For more on the typology of verb borrowing, see Wichmann & Wohlgemuth (2008), who also show that the kind of borrowing can serve as an indicator of the degree of bilingualism between the languages. Wichmann & Wohlgemuth propose a hierarchy of borrowing strategies where incorporating a new verb by ways of a construction with a form of the native 'do' (or some other so-called light verbs like 'be' or 'make', often in conjunction with a form of 'do') is at the top of the hierarchy and indicates a high degree of bilingualism, while borrowing a whole paradigm (i.e. the verb including its affixes, a rare strategy) is at the bottom of the hierarchy and indicates a low degree of bilingualism. The two intermediate stages are what they call 'Indirect insertion' where an affix is needed in the recipient language to integrate the borrowed verb and 'Direct insertion' where the verb stem is simply borrowed and integrated in the grammar of the recipient language without any further accommodation. For more details, see Wichmann & Wohlgemuth (2008).

between languages. Purépecha, for example, has, among other markers, borrowed the Spanish *pues* 'thus, then, well' and *bueno* 'well', pronounced *pwes* or *pos* and *wenu* respectively (Chamoreau 2007: 475).

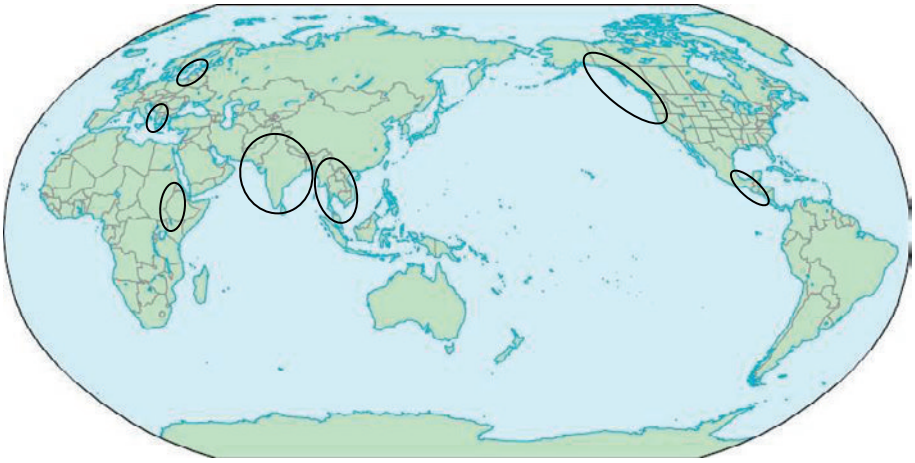
13.2.3 Linguistic areas

Consistent contact between languages may give rise to a **linguistic area** (also termed ***Sprachbund***, **diffusion area**, **convergence area** and **adstratum relationship**), which essentially means that a group of not closely related languages in a delimited area share a combination of structural traits. **Areal features** are thus a bundle of features consistently found in the languages of a particular area. It is important to keep in mind that different areal features may have different origins and have diffused to different degrees within a linguistic area. For a very accessible overview of areal linguistics, see Campbell (2009).

A high number of linguistic areas of different sizes have been identified across the world. Below, I list a few of the most well-known ones together with a few examples of their distinctive traits, largely following the salient features listed in Campbell (2009). It is beyond the scope of this section to discuss all the various linguistic areas that have been identified world-wide. Apart from the linguistic areas discussed below, Europe and 'Standard Average European' (see Haspelmath 1998 and subsequent), the Southeast of the USA (see Campbell 1997), the Amazon (see Aikhenvald & Dixon 1998 with further references) in South America, the Central Andes (see Büttner 1983) in Western South America, and many more are well-defined linguistic areas. Furthermore, I am confining myself to only five distinctive traits per linguistic area, which for none of the areas constitutes a complete list of salient features. Note that these features do not necessarily need to be typologically rare; it is the fact that they occur in a cluster that is of interest. It is also important to keep in mind that every feature of a linguistic area does not necessarily occur in every single language of a linguistic area. For most of the linguistic areas the number of languages included is too high to provide a list of the individual names. Map 13.1 gives a rough indication of where the linguistic areas discussed below are located. I stress again that this is only a small selection of the known linguistic areas world-wide.

13.2.3.1 *The Balkans*

This was the first *Sprachbund* to be recognized by linguists and is perhaps the best-known and discussed linguistic area. It consists mainly of Indo-European languages of different genera, although Macedonian Turkish constitutes an exception. The languages usually included in this area are Greek, Albanian (Albanian: Albania), Serbian (Slavic: Serbia), Croatian, Bulgarian (Slavic: Bulgaria), Macedonian (Slavic: Macedonia), Romanian (Indo-European (Romance): Romania), Bugurdži Romani (Indo-European (Indic): Serbia) and Macedonian Turkish.



Map 13.1 A rough indication of the linguistic areas discussed below: circum-Baltic, Balkan, Ethio-Eritrean, South Asian, mainland Southeast Asian, Northwest Pacific coast and Meso-American linguistic areas. These represent only a handful of the identified areas world-wide.

Below are a few of the most distinguishing features of the Balkan area. For more details and further references see, for example, Joseph (1992).

a. a central vowel /ɨ/ or /ə/
b. postposed articles
c. an analytic (or periphrastic) future
d. absence of infinitives (instead expressed 'I want *that I go*')
e. using a pronoun copy to double mark human objects as in (374):

Greek (Indo-European (Greek): Greece)

(374)	**ton**	vlépo	**ton**	jáni	
	3SG.ACC	see.PRES.1SG	3SG.ACC	PN	
	'I see Jani.' (lit. him I see him Jani)				(Campbell 2009: 62)

13.2.3.2 *The circum-Baltic*

The circum-Baltic area includes at a minimum Estonian, Liv (Uralic (Finnic): Latvia), Lithuanian (Indo-European (Baltic): Lithuania), Latvian, Baltic German (Indo-European (Germanic): Germany), but other languages have also been included, such as Finnish, Swedish, Karelian (Uralic (Finnic): Russia), Karaim (Altaic (Turkic): Lithuania), Baltic Yiddish, and many more. In other words, the languages around the eastern Baltic Sea area.

Below are a few of the most distinguishing features of the circum-Baltic area. For more details and further references see, for example, Koptjevskaja-Tamm & Wälchi (2001).

a. first syllable stress
b. partitive case or partitive constructions
c. evidential mood
d. adjectives agreeing in number with the head noun
e. nominative case for direct objects in constructions that lack overt subjects

13.2.3.3 *Ethio-Eritrea*

The Ethio-Eritrean area (traditionally called the Ethiopian area) includes some Nilo-Saharan languages such as Anywa (Nilotic: Sudan), Gumuz (Gumuz: Ethiopia), as well as a number of Cushitic languages (among others Qafar (Eastern Cushitic: Ethiopia), Awngi (Central Cushitic: Ethiopia), Beja, Sidaama (Eastern Cushitic: Ethiopia) and Somali), Semitic languages (among others Amharic, Ge'ez and Tigre in Ethiopia as well as Tigrinya in Eritrea) and Omotic languages (among others Kefa (South Omotic), Wolaytta (North Omotic) and Yemsa (North Omotic), all in Ethiopia), all of which belong to the Afro-Asiatic family. It is a large area that essentially comprises the Ethiopian highlands and stretches into Eritrea as well as northern Kenya.

Below are a few of the most distinguishing features of the Ethio-Eritrean area. For more details and further references see, for example, Ferguson (1976) and Zaborski (1991). For a discussion why the Ethiopian area should not be considered one linguistic area but rather a cluster of smaller linguistic areas, see Tosco (2000).

a. the pharyngeal fricatives /ħ/ and /ʕ/
b. negative copula
c. a postposed quotative
d. gender distinction in second and third person pronouns
e. a general number for the base form and overt marking of the singular

13.2.3.4 *South Asia*

The South Asian area is a vast area of languages spoken on the South Asian subcontinent (Pakistan, India, Nepal, Bangladesh, and Bhutan) as well as the island of Sri Lanka. It includes Indo-European, Dravidian, Tibeto-Burman and Munda languages.

Below are a few of the most distinguishing features of the South Asian area. For more details and further references see, for example, Emeneau (1956) and Thomason (2001: 114ff).

a. retroflex consonants
b. absence of prefixes
c. morphological causatives
d. absence of the lexical verb 'to have'
e. echo constructions with the meaning 'and so on' or 'and the like'

13.2.3.5 *Mainland Southeast Asia*

The Southeast Asian, or more precisely, the mainland Southeast Asian linguistic area is also a huge area including Vietnam, Laos, Cambodia and Thailand, but which also stretches into Myanmar (Burma), southern China and Northern mainland Malaysia. It includes Austronesian, Hmong-Mien, Mon-Khmer, Sinitic (a branch of Sino-Tibetan) and Tai (a branch of Tai-Kadai) languages.

Below are a few of the most distinguishing features of the mainland Southeast Asian area. For more details and further references see, for example, Enfield (2001 and 2005).

a. lexical tone
b. serial verb constructions
c. classifier constructions
d. absence of cases
e. isolating morphology

13.2.3.6 *The Northwest Pacific Coast*

The northwest Pacific Coast linguistic area stretches from northern California all the way to Alaska. This long linguistic corridor includes languages from the Chimakuan, Na-Dene, Oregon Coast, Penutian, Salishan and Wakashan families.

Below are a few of the most distinguishing features of the Northwest Pacific Coast area. For more details and further references see, for example, Campbell (1997: 332ff).

a. rich consonant systems (with multiple laterals and uvulars as well as rich glottalization) but limited vowel systems
b. several kinds of reduplication
c. sentence initial negation
d. alienable/inalianable possession
e. evidentiality

13.2.3.7 *Meso-America*

The Meso-American linguistic area stretches from central Mexico, over Belize and Guatemala to El Salvador, southern Honduras and the Pacific coast of Nicaragua and northern Costa Rica. It contains a huge amount of languages from the Huave, Nahua

(a branch of Uto-Aztecan), Mayan, Mixe-Zoque, Oto-Manguean, Tarascan, Totonacan, Xincan and Tequistlatec families, as well as the isolate Cuitlatec (Mexico).

The following five traits are shared by all languages in the Meso-American area. For more details on other traits, shared by many but not all of the languages in the area, and for further references see, for example, Campbell et al. (1986).

a. a vigesimal numeral system
b. possession of the type *his-dog the man* 'the man's dog', as in Example (375):

Quiché (Mayan (Mayan): Guatemala)

(375) u-ɕ'i:ʔ le: ačih
his-dog the man
'the man's dog' (Campbell et al. 1986: 545)

c. relational nouns for locative expressions (such as *at.my.back* for 'behind me')
d. VO word order (despite being surrounded by OV languages)
e. a large set of lexical calques (loan translations) not found outside the linguistic area, such as 'child of hand' for 'finger', 'mother of hand' for 'thumb', 'mouth of house' for 'door', 'water-die' for 'thirst' and many more (Campbell et al. 1986: 553)

13.3 Language change in sign languages

Just like spoken languages, sign languages constantly change. Change may be due to either internal or external factors. In other words, processes of grammaticalization are identifiable in sign languages, just as they are in spoken languages. One major difference between spoken and signed languages is that sign languages are almost universally in intense contact situations: most signers live in a world entirely dominated by spoken language(s).

In many cases grammaticalization in sign languages follows the same or similar kinds of paths as grammaticalization in spoken languages, but there are some paths that seem to be specific for signed languages. For an extremely accessible and thorough investigation of grammaticalization paths in sign languages and in how far they are parallel to those in spoken languages, see Pfau & Steinbach (2006a), which this section relies a great deal on.

Noun > Pronoun. In both DGS and NGT the indefinite pronoun ('one, someone') < ONE^PERSON (with a reduced form of the numeral one) (Pfau & Steinbach 2006a: 31).

Verb > Auxiliary. In many sign languages the verb 'finish' has developed into an aspect marker; for example, in LIS FATTO 'finish/done' > perfective aspect marker

(Pfau & Steinbach 2006a), while in ISL FINISH > completive aspect marker (Meir 1999). In ASL the future tense auxiliary < GO.TO (from the Old French Sign Language (OFSL: France) PARTIR 'leave') and the modal auxiliary 'must' < OWE (from OFSL IL.FAUT 'must').[209] The latter path is particularly interesting, deriving ultimately from the gesture of monetary debt, desemanticizing to a general meaning of 'owing' and extending to a meaning of 'obligation'. For more on ASL future tense and modal markers, see Janzen & Schaffer (2002: 203ff), from where I have the above examples.

Adjective > Auxiliary. The ASL modal auxiliary 'can' < STRONG (ultimately from OFSL POUVOIR 'be.strong'; Janzen & Shaffer 2002: 207ff). While many modal auxiliaries denoting ability in spoken languages also ultimately derive from lexical items denoting physical strength of capacity, they tend to grammaticalize via lexical verbs and not adjectives (cf. Bybee et al. 1994).

Adverb > Auxiliary. In DGS, NGT and ISL ALREADY > perfective aspect marker (Pfau & Steinbach 2006a).

Noun > Auxiliary. This grammaticalization path is not known for spoken languages. In DGS the so-called agreement auxiliaries, which are auxiliaries used with non-agreeing (or non-directional) verbs (cf. 5.3), are derived from the noun PERSON (Pfau & Steinbach 2006a: 28ff).

Noun > Complementizer. In DGS, for example, a phonologically reduced form of the noun REASON functions as the complementizer 'because' (Pfau & Steinbach 2006a: 35f).

The ultimate origins of the future tense and modal auxiliaries mentioned above seem to have been gestures, though their grammaticalization paths all went through a lexical stage. A phenomenon specific to sign languages is the direct grammaticalization of gestures: here the path goes straight from an iconic item to a grammatical item without the intermediate step of fist having developed into a lexical item.[210] Below are a few examples (all from Pfau & Steinbach 2006a: 52ff unless otherwise indicated).

---

209. "The historical link between OFSL and ASL is due to the fact that Thomas Gallaudet who founded the first school for the deaf [in USA] (the American Asylum in Hartford, CT) in 1816 went to Paris to learn signs and methods of instructing deaf children. Gallaudet returned to America with Laurent Clerc, himself a deaf graduate of the Paris school. Woodward (1978) suggests that what is now known as ASL constitutes, in large part, a mix of the lexicon and some elements of the grammar of OLSF with an indigenous sign language used in the North-Eastern part of the USA at that time" (Pfau & Steinbach 2006a: 18). This essentially makes ASL an OLSF-lexified creole. For more on the history of ASL, see Lane (1984).

210. But cf. the discussion on demonstratives in Diessel (1999a) mentioned above. This might possibly constitute an example of spoken language functional items entering the linguistic system without the intermediate lexical stage.

Gesture (manual) > Classifier. The handle classifiers that are so common across sign languages do not generally originate from nouns but from gestures. The FLOWER-classifier in DGS, for example, resembles the letter F while the noun flower is signed with the fingers extending from a closed hand in an upwards movement. In other words, there is no resemblance between the classifier and the noun.

Gesture (manual) > Question particle. The "palm-up" gesture that many speakers use as a gesture when uttering a question has grammaticalized to question particles in many sign languages. In most sign languages it is used as a polar question particle, but in IPSL the same gesture covers both polar and content questions.

Gesture (non-manual) > Negation. The non-manual side-to-side headshake has grammaticalized in many sign languages into a sentential negator.

Gesture (non-manual) > Polar question marker > Topic marker. The non-manual "communicative questioning gesture" (Janzen & Shaffer 2002: 212) of raised eyebrows has developed into an obligatory polar question particle in many sign languages. This gesture may then extend to also become a topic marker, as it has in, for example, ASL.

Sign languages are also susceptible to contact-induced change, just as spoken languages are. The difference between spoken and signed languages, however, is, as mentioned, the fact that virtually all sign languages are minority languages in a speaker dominated society. The absolute majority of sign language users must be able to function in the spoken language of the society they live in; until recently many societies even forced deaf and hard of hearing people to learn to speak the spoken language of the society while the natural sign language was repressed. Most users of sign language are thus bilingual (effectively bimodally bilingual, cf. Emmorey et al. 2005).

Linguistic contact with sign languages may be of two types: between different sign languages, or between signed and spoken languages. Code-switching occurs in sign languages as in spoken languages, and may either be between two sign languages or between a signed and a spoken language. The latter case may be better termed as **code-blending** (Emmorey et al. 2005), since they typically involve simultaneous signing and speaking. An example of a code-blend between spoken English and ASL is the following:

Spoken English/ASL code-blend

(376) I [don't] [think] he would [really] [live]
NEG THINK REALLY LIVE (Emmorey et al. 2005: 666)

In (376) parts of the spoken utterance (indicated by square brackets) are accompanied with signs (glossed in the conventional manner).

Borrowings may occur between sign languages as well as between spoken and signed languages. Examples of borrowings between sign languages are country names,

which are increasingly signed using the native sign of the language in question (the CSL sign for CHINA is taken over in other sign languages, and so on). Other examples of lexical borrowings are the dialects of BSL found in Glasgow, Liverpool and London where the Catholic Deaf community have borrowed from TCE, for example the signs of YEAR and BABY.DAUGHTER (Sutton-Spence & Woll 1999: 220).

Examples of borrowings from spoken languages to signed languages are calques such as ASL BOY + FRIEND or HOME + WORK from spoken English *boyfriend* and *homework* (Valli et al. 2005: 187). This essentially constitutes a case of pattern replication. Another example is the ASL past auxiliary < the spoken English prefix *ex* (Carol Neidle, p.c. to Anthony Grant). An interesting case of matter borrowing from ASL to spoken English has been reported by Valli et al. (2005: 187), whereby the non-manual mouthing *cha* which accompanies signs for 'large pile of papers' or 'thick book' has been heard in such utterances as *I have **cha** homework*.

Borrowings are not necessarily only of lexical nature. Many of the grammaticalized gestures are cases of contact-induced change. That is, the gestures are typically found in the spoken languages surrounding the sign languages in question, and have either been transferred into the sign languages or have enhanced an already existing tendency.

13.4 Summary

All human languages change both due to internal and external factors. A major internal process of language change is grammaticalization. Language external change is mainly due to contact between languages.

The process of grammaticalization typically involves desemanticization, where the semantic content of a lexical item is bleached, which allows the item to be extended in its contextual use. This increases the frequency of use of the item, which in turn leads to decategorialization as well as morphological and phonological erosion. Paths of grammaticalization usually involve lexical items that grammaticalize into functional items. The starting point tends to be content words that grammaticalize to function words, which may grammaticalize into clitics, which in turn may grammaticalize into affixes. Eventually the item may erode to the point of loss. Grammaticalization is overwhelmingly unidirectional.

Contact-induced change typically comes about due to bi- or multilingualism and code-switching. There are different types of contact situations, the main difference boiling down to dominance. In a balanced contact situation the languages in question are on a roughly equal footing. In a displacive contact situation one language dominates over the other. In cases of extreme dominance speakers may end up shifting to the dominating language, which ultimately may result in the extinction of the dominated language.

In principle, any feature may be borrowed from one language into another, but not all features are equally prone to diffusion. Borrowings are generally of two basic types, matter borrowing and pattern replication. Language contact may also induce grammaticalization.

Language contact may give rise to linguistic areas, where a geographically delineated area contains several not closely related languages sharing a cluster of linguistic features that are not found among the languages outside the linguistic area.

Sign languages are also subject to constant change, both by way of internal and external factors. In general, grammaticalization paths in sign languages tend to be similar to those in spoken languages. A modality-specific path for sign languages is the grammaticalization of gestures to grammatical categories without the intermediate lexical stage.

Almost all sign languages are in a situation where they are dominated by the spoken languages of the society in question. This typically leads to bimodal bilingualism which in turn may lead to bimodal code-blending. Sign languages borrow from each other and from spoken languages. Contact-induced grammaticalization of function categories by way of grammaticalization of gestures is attested in most known sign languages.

13.5 Keywords

areal feature
"bleaching"
borrowing
cline
code-switching
decategorialization
diffusion
erosion
extension
grammaticalization
language contact
language shift
linguistic areas
unidirectionality

13.6 Exercises

1. In what way does grammaticalization explain processes of analytic constructions becoming synthetic?
2. In what way are demonstratives (possibly) exceptional in terms of grammaticalization?
3. What is the difference between **balanced** and **displacive** language contact?
4. Define ***Sprachbund***.
5. Is the following statement true or false? Motivate your answer.

   The mode of communication might affect grammaticalization paths.

Appendix 1

Selected sample of sites

This is not an exhaustive list of what is available on the internet. The sites are listed in alphabetical order.

Atlas of Pidgin and Creole Language Structures
(http://lingweb.eva.mpg.de/apics) Note that this is still a temporary address.

A project gathering comparable data on lexical and grammatical structures of 76 different pidgins, creoles and mixed languages.

Auslan Signbank
(http://www.auslan.org.au/)

A resource site for Auslan featuring dictionaries, linguistic examples and general information about the language and community.

Automated Similarity Judgement Program
(http://wwwstaff.eva.mpg.de/~wichmann/ASJPHomePage.htm)

A project aiming at a computerized lexicostatistical analysis and classification of the world's languages.

AUTOTYP
(http://www.uni-leipzig.de/~autotyp/)

A large scale project aiming at adding linguistic types as they come to linguists' attention for use in both quantitative and qualitative typological research.

British Sign Language Corpus Project
(http://www.bslcorpusproject.org/)

An online corpus of BSL, including regional varieties and sociobiographical information on the signers.

Creolica
(http://www.creolica.net/)

An open-access journal publishing articles relating to the study of creole languages in particular and typological issues in general.

Das Grammatische Raritätenkabinett (The Grammatical Rarity Collection)
(http://typo.uni-konstanz.de/rara/intro/index.php)

A collection of linguistic rarities assembled by Frans Plank, including further references, available as a searchable archive.

Directory of Open Access Journals
(http://www.doaj.org/doaj?func=home&uiLanguage=en)

A directory listing free, full text and quality controlled scholarly journals in most subjects.

DGS-Corpus
(http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html)

An online corpus and dictionary of DGS, including regional varieties.

Ethnologue
(http://www.ethnologue.com/home.asp)

An encyclopaedic reference work cataloguing all of the world's known living languages, including macro data such as number of speakers, location of language, language use and some further references.

Language Description Heritage
(http://ldh.livingsources.org/)

An open access digital library aiming to provide easy access to descriptive material about the world's languages.

Language Documentation & Conservation
(http://nflrc.hawaii.edu/ldc/)

A blind peer reviewed open-access journal publishing articles on topics related to language documentation and conservation.

Linguistic Discovery
(http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/2/xmlpage/1/issue)

A blind peer reviewed open-access typologically oriented journal, focussing especially on issues related to primary data as well as lesser studied languages and phenomena.

Living Tongues – Institute for Endangered Languages
(http://www.livingtongues.org/)

A project promoting the documentation, preservation, maintenance and, if possible, revitalization of endangered languages through the combined efforts of linguists and the affected communities.

Roger Blench
(http://www.rogerblench.info/RBOP.htm)

A site containing masses of information related to the work of Roger Blench, including a vast number of shorter linguistic fieldwork descriptions and datasheets for languages from all over the world, predominantly Africa.

Romani Morpho-Syntax Database
(http://romani.humanities.manchester.ac.uk/rms/)

Database of over 150 varieties of Romani, focussed on morpho-syntax but also containing phonological and lexical material; all based on fieldwork and including sound files.

Sign Language Linguistics Society
(http://www.slls.eu/index2.php5)

A resource site with the aim of promoting sign language research. Contains a wealth of information on events and publications in the field.

Summer Institute of Linguistics
(http://www.sil.org/)

An organization with Christian overtones engaging in massive efforts of translation, description, training and production of language materials, while at the same time providing the linguistic community with a host of data, literature and various essential tools (e.g. fonts and transcription programs, among other things).

Stresstyp
(http://stresstyp.leidenuniv.nl/)

A searchable database of information on the metrical systems of over 500 languages.

Surrey Morphology Group Databases
(http://www2.surrey.ac.uk/english/smg/)

A collection of databases collecting information on grammatical categories and morphological strategies in a broad sample of languages, as well as addressing terminological issues by formulating canonical definitions of them.

Syntactic Structures of the World's Languages
(http://sswl.railsplayground.net/)

An open-ended searchable database which allows the user to look up the collected properties of a particular language as well how specific properties relate across languages.

The Corpus NGT
(http://www.ru.nl/corpusngten/)

An online corpus of NGT.

The Electronic World Atlas of Varieties of English
(http://www.ewave-atlas.org/)

An open access online database on morphosyntactic variation in 48 varieties of English and 26 English-lexified pidgin and creole languages worldwide.

The Linguistics Research Center
(http://www.utexas.edu/cola/centers/lrc/)

A site chiefly devoted to ancient Indo-European languages and cultures, containing access to literature, language lessons, lexica (also etymological), maps and old texts as well as bibliographies for further references.

The Rosetta Project
(http://rosettaproject.org/)

An international collaboration of linguists and native speakers aiming at producing a publicly accessible digital library of human languages for long-term archiving, all of which are described according to a uniform format, including a longer text, vocabulary and description, as well as further references for each language.

The UCLA Phonological Segment Inventory Database
(http://web.phonetik.uni-frankfurt.de/upsid.html)

An open access database on the phonological systems of several hundred languages, searchable through the web interface developed by Henning Reeth (which is the link given above).

The Universals Archive
(http://typo.uni-konstanz.de/archive/intro/)

A site aiming at collecting and testing linguistic universals mentioned in the literature, especially implicational universals, available as a searchable archive.

The World Atlas of Language Structures Online
(http://wals.info/)

An open access searchable database covering structural information on over 2500 languages worldwide, including classifications and location of the languages, as well as links to information on the languages on other sites as well as the references used for the individual languages.

The World Loanword Database
(http://wold.livingsources.org/)

A collection of vocabularies from 41 languages for the use of investigations in loanword typology.

Typological Database System Project
(http://languagelink.let.uu.nl/tds/index.html)

A collection of typological databases pulled together and integrated to one searchable unit covering, in total, over 1000 languages.

UCLA Phonetics Lab
(http://archive.phonetics.ucla.edu/)

An open access database of recordings, including phonetic transcriptions as well as field notes, of several hundred languages from around the world

XTone
(http://xtone.linguistics.berkeley.edu/)

A forum meant for researchers interested in contributing and accessing information on tone systems in the languages of the world.

Appendix 2

Languages cited in this book

The following lists the languages that were cited in the text of the book. It does not include the languages of the various feature maps that were not mentioned in the text. The language name is as it appears in the text; if this differs from the name listed in Ethnologue, the latter is given in parentheses. Thus "Acoma (Keres, Western)" means that *Acoma* is the language name used in the text of this book, while the same language is called *Keres, Western* in Ethnologue. The ISO 639-3 code for each language is given where available. For the sign languages the English names are given. The locations are given as country names and should be seen as rough approximations. The figures for the number of speakers are taken from Ethnologue unless otherwise specified, and refer to the total for all countries given a particular language. The figures for the number of speakers also refer to the whole language unless otherwise specified, i.e. are not broken down into numbers for specific dialects. The year indicates when the figure for the number of speakers was given; "NA" means that no estimate of number of speakers is available and "(n.y.)" means that no year was given for the estimate in question.

Language	ISO	Family	Genus	Location	Speakers	Page
!Xóõ	nmn	Khoisan	Southern Khoisan	Botswana	4200 (2002)	72, 384, 393
Aari	aiw	Afro-Asiatic	South Omotic	Ethiopia	155,000 (1994)	14, 31, 114, 141
Abkhaz	abk	North-West Caucasian	North-West Caucasian	Georgia	117,350 (1993)	72, 344, 359, 361
Acholi	ach	Nilo-Saharan	Nilotic	Uganda	791,796 (1991)	384, 407–408
Acoma (Keres, Western)	kjq	Keresan	Keresan	USA	3391 (1980)	114, 135, 306, 326
Adamorobe Sign Language	ads	Sign Language	Sign Language	Ghana	3400 (2003)	5, 14, 30, 38, 58, 60, 86, 276, 303, 306, 342
Afrikaans	afr	Indo-European	Germanic	South Africa	4,934,950 (2006)	276, 286, 344, 372
Ahus, Hus (Andra-Hus)	anx	Austronesian	Oceanic (Eastern Malayo-Polynesian)	Papua New Guinea	1,310 (2000)	
Aikaná (Tubarão)	tba	Isolate	Isolate	Brazil	180 (2005)	60, 76, 228, 259
Ainu	ain	Isolate	Isolate	Japan	15 (1996)	14, 22, 32–33, 114, 131, 276, 285, 344, 354
Aiome	aki	Lower Sepik-Ramu	Annaberg	Papua New Guinea	750 (1981)	114, 143
Albanian (Albanian, Tosk)	als	Indo-European	Albanian	Albania, Serbia, Montenegro	3,035,000 (1989)	384, 411
Alemannic, High (German, Swiss)	gsw	Indo-European	Germanic	Switzerland	6,469,000 (2000)	38, 53
Al-Sayyid Bedouin Sign Language	---	Sign Language	Sign Language	Israel	ca 100 (2008) (Aronoff et al. 2008)	60, 86, 228, 273, 304
Alutor	alr	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan	Russia (Siberia)	150 (2000)	88, 109
Ambon Malay (Malay, Ambonese)	abs	Creole	Malay-lexified	Indonesia	ca 200,000 (Paauw 2013)	114, 174, 344, 372
Amele	aey	Trans-New Guinea	Madang	Papua-New Guinea	5,300 (1987)	306, 327, 337
American Sign Language	ase	Sign Language	Sign Language	USA	100,000–500,000 (1986)	14, 28–29, 60, 84, 86, 88, 110–112, 114, 151, 154, 190, 192, 224–226, 228, 273, 306, 340, 344, 381, 384, 416–418

Language	ISO	Family	Genus	Location	Speakers	Page
Amharic	amh	Afro-Asiatic	Semitic	Ethiopia	17,528,500 (1994)	384, 413
Amis	ami	Austronesian	Paiwanic	Taiwan	138,000 (2002	306, 331
Angas (Ngas)	anc	Afro-Asiatic	West Chadic	Nigeria	400,000 (1998)	154, 170
Angolar	aoa	Creole	Portuguese-lexified	São Tomé e Príncipe	5,000 (Maurer 2013a)	344, 366
Anindilyakwa	aoi	Australian	Anindilyakwa	Australia	1,240 (1996)	114, 142, 154, 180
Anjam	boj	Trans-New Guinea	Madang	Papua New Guinea	2,020 (2003)	192, 216
Anywa (Anuak)	anu	Nilo-Saharan	Nilotic	Sudan, Ethiopia	97,600 (1991)	384, 413
Apurinã	apu	Arawakan	Arawakan	Brazil	extinct? Ethnic population: 4,087 (2003)	192, 221
Arabic, Egyptian	arz	Afro-Asiatic	Semitic	Egypt	53,990,000 (2006)	88, 98
Arabic, Standard	arb	Afro-Asiatic	Semitic	Northern Africa, Middle East	206,000,000 (1999)	60, 73, 96, 114, 131, 384, 410
Araona	aro	Tacanan	Tacanan	Bolivia	81 (2000)	228, 275
Argentine Sign Language	aed	Sign Language	Sign Language	Argentina	NA	14, 29, 276, 303–304
Arrernte, Mparntwe (Arrernte, Eastern)	aer	Australian	Pama-Nyungan	Australia	3820 (1996)	60, 72–73
Assamese	asm	Indo-European	Indic	India	16,818,750 (2000)	276, 291
Asuriní (Asurini, Tocantins)	asu	Tupian	Tupi-Guaraní	Brazil	300 (2001)	276, 286
Australian Sign Language	asf	Sign Language	Sign Language	Australia	102,000 (n.y.)	14, 28–29, 156, 190, 192, 225, 306, 340, 421
Austrian Sign Language	asq	Sign Language	Sign Language	Austria	NA	114, 151, 154, 190, 276, 304
Awa Pit (Awa-Cuaiquer)	kwi	Barbacoan	Barbacoan	Colombia	22,000 (1986)	228, 250
Awngi	awn	Afro-Asiatic	Central Cushitic	Ethiopia	500,000 (2007)	384, 413
Aymara (Aymara, Central)	ayr	Aymaran	Aymaran	Bolivia	2,262,900 (1987)	228, 240

Language	ISO	Family	Genus	Location	Speakers	Page
Babanki	bbk	Niger-Congo	Bantoid (Benue-Congo)	Cameroon	22,500 (2000)	75
Babungo (Vengo)	bav	Niger-Congo	Bantoid	Cameroon	27,000 (2008)	228, 251, 306, 310, 327
Bajau, Sama (Bajau, Indonesian)	bdl	Austronesian	Sama-Bajau	Indonesia	150,000 (2007)	276, 289
Baka	bkc	Niger-Congo	Ubangi	Cameroon	43,200 (2007)	306, 327
Balanta (Balanta-Ganja)	bjt	Niger-Congo	Northern Atlantic	Guinea Bissau, Senegal	82,800 (2006)	5, 114, 127
Bali-Vitu (Uneapa)	bbn	Austronesian	Oceanic	Papua New Guinea	10,000 (1998)	228, 247
Bambara (Bamanankan)	bam	Niger-Congo	Western Mande	Mali	2,772,340 (1995)	5, 60, 69, 114, 127
Ban Khor Sign Language	bfk	Sign Language	Sign Language	Thailand	NA	38, 58
Bandjalang	bdy	Australian	Pama-Nyungan	Australia	10 (1983)	60, 65
Banyun (Bainouk-Gunyaamolo)	bcz	Niger-Congo	Northern Atlantic	Senegal	32,520 (2006)	154, 162
Barai	bbb	Trans-New Guinea	Koiarian	Papua New Guinea	800 (2003)	306, 331
Bashkir (Bashkort)	bak	Altaic	Turkic	Russia	1,451,340 (2002)	60, 83
Basque	eus	Isolate	Isolate	Spain, France	658,960 (1991)	14, 22, 228, 241, 268, 384, 395, 405, 408
Basque, Goizueta	eus	Isolate	Isolate	Spain	658,960 (1990)	60, 80
Batavia Creole	---	Creole	Portuguese-lexified	Indonesia	extinct (Maurer 2013b)	60, 82, 344, 367, 372
Bauchi Guda	bsf	Niger-Congo	Kainji	Nigeria	10–20,000 (Harley & Blench fc)	60, 75–76
Bayso (Baiso)	bsw	Afro-Asiatic	Eastern Cushitic	Ethiopia	1,010 (1995)	154, 159, 161
Begak-Ida'an (Ida'an, Begak)	dbj	Austronesian	Northwest Malayo-Polynesian	Malaysia	1,500 (2000)	114, 134
Beja (Bedawiyet)	bej	Afro-Asiatic	Beja	Eritrea, Sudan	1,186,000 (1982)	88, 98, 384, 413
Belizean Creole (Belize Kriol English)	bzj	Creole	English-lexified	Belize	150,000 (Escure 2013)	114, 139

Language	ISO	Family	Genus	Location	Speakers	Page
Bemba	bmy	Niger-Congo	Bantoid	DR Congo	296,000 (2000)	114, 129
Berawan, Central	zbc	Austronesian	Northwest Malayo-Polynesian	Malaysia	710 (2007)	60, 79
Berber, Figuig	---	Afro-Asiatic	Berber	Morocco	NA	38, 41
Berber, Middle Atlas (Tamazight, Central Atlas)	tzm	Afro-Asiatic	Berber	Morocco	3,150,000 (1998)	88, 98, 228, 240
Bété (Bété, Daloa)	bev	Niger-Congo	Kru	Ivory Coast	130,000 (1993)	60, 74
Biloxi	bll	Siouan	Siouan	USA	extinct	344, 360–361
Bislama	bis	Creole	English-lexified	Vanuatu	"widely spoken among total population of ca 218,000" (Meyerhoff 2013)	114, 139, 228, 249
Bororo (Borôro)	bor	Macro-Ge	Bororo	Brazil	1,020 (1997)	88, 106
Brazilian Sign Language	bzs	Sign Language	Sign Language	Brazil	NA	14, 29
British Sign Language	bfi	Sign Language	Sign Language	UK	40,000 (1984)	xxii, 7, 14, 28–30, 192, 224–225, 276, 303, 344, 379, 384, 418, 421
Bulgarian	bul	Indo-European	Slavic	Bulgaria	9,097,220 (1986)	384, 411
Buma (Teanu)	tkw	Austronesian	Oceanic	Solomon Islands	520 (2007)	276, 294
Burmese	mya	Sino-Tibetan	Burmese-Lolo	Myanmar	32,319,700 (2000)	60, 65, 344, 351, 354
Burunge	bds	Afro-Asiatic	Southern Cushitic	Tanzania	13,000 (2002)	154, 170, 344, 355
Canela-Krahô	ram, xra	Macro-Ge	Ge-Kaingang	Brazil	3,800 (2001)	228, 243
Cantonese (Chinese, Yue)	yue	Sino-Tibetan	Chinese	China	55,541,660 (1984)	114, 138, 306, 331, 384, 409
Cape Verdean Creole of Brava (Kabuverdianu)	kea	Creole	Portuguese-lexified	Cape Verde Islands	6043 (2010; Baptista 2013)	154, 171
Cape Verdean Creole of Santiago (Kabuverdianu, Sotavento)	kea	Creole	Portuguese-lexified	Cape Verde Islands	450.000 (Lang 2013)	154, 171

Language	ISO	Family	Genus	Location	Speakers	Page
Catalan Sign Language	csc	Sign Languages	Sign Languages	Spain	18,000 (1994)	xx, 114, 152
Cayuga	cay	Iroquoian	Northern Iroquoian	USA	60 (2002)	276, 301
Cayuvava (Cayubaba)	cyb	Isolate	Isolate	Bolivia	extinct	154, 179
Cèmuhî	cam	Austronesian	Oceanic	New Caledonia	2,050 (1996)	276, 285
Chabacano, Caviteño (Chavacano, Caviteño)	cbk	Creole	Spanish-lexified	Philippines	4,000 (Sippola 2013a)	88, 103, 276, 286, 344, 372
Chabacano, Ermiteño (Chavacano, Ermiteño)	cbk	Creole	Spanish-lexified	Philippines	extinct	88, 103
Chabacano, Ternateño (Chavacano, Ternateño)	cbk	Creole	Spanish-lexified	Philippines	less than 3,000 (Sippola 2013b)	88, 103, 276, 286
Chabacano, Zamboanga (Chavacano, Zamboanga)	cbk	Creole	Spanish-lexified	Philippines	300,000 (Steinkrüger 2013)	114, 141, 276, 286, 344, 372
Chácobo	cao	Panoan	Panoan	Bolivia	550 (2000)	192, 199
Chamicuro	ccc	Arawakan	Arawakan	Peru	2 (2000)	192, 194, 196
Chamorro	cha	Austronesian	Chamorro	Guam	92,700 (1991)	60, 78, 228, 268
Chantyal	chx	Sino-Tibetan	Bodic	Nepal	2,000 (1997)	228, 248
Chechen	che	Nakh-Daghestanian	Nakh	Russia	1,341,000 (2002)	154, 183
Chemehuevi (Ute-Southern Paiute)	ute	Uto-Aztecan	Numic	USA	13 (1994) specifically Chemehuevi dialect	228, 247–248
Chepang	cdm	Sino-Tibetan	Bodic	Nepal	36,800 (2001)	228, 250
Chichewa (Nyanja)	nya	Niger-Congo	Bantoid	Malawi, Mozambique, Zambia, Zimbabwe	8,659,700 (2001)	88, 97–98, 109, 344, 348
Chinantec, Quiotepe	chq	Oto-Manguean	Chinantecan	Mexico	8,000 (1998)	60, 76
Chinese Pidgin English	cpi	Pidgin	English-lexified	(Chinese S coast)	"rememberers only" (Matthews & Li 2013)	154, 163, 174

Language	ISO	Family	Genus	Location	Speakers	Page
Chinese Pidgin Russian	---	Pidgin	Russian-lexified	(Russia/China border)	extinct, but ca 50 semi-speakers (Perekhvalskaya 2013)	114, 154, 163, 174, 306, 344, 367
Chinese Sign Language	csl	Sign Language	Sign Language	China	NA	14, 29, 276, 304, 344, 380, 384, 418
Chinese, Classical	---	Sino-Tibetan	Chinese	China	extinct	154, 159
Chintang (Chhintange)	ctn	Sino-Tibetan	Bodic	Nepal	1,500 (2003)	88, 92
Chinuk Wawa (Chinook Wawa)	chn	Pidgin	Chinook-lexified	Canada, USA	1 (Grant 2013)	114, 140, 154, 163
Choctaw	cho	Muskogean	Muskogean	USA	11,400 (1998)	228, 254, 276, 291
Chol (Chol, Tumbalá)	ctu	Mayan	Mayan	Mexico	90,000 (1992)	228, 257
Chontal Maya (Chontal, Tabasco)	chf	Mayan	Mayan	Mexico	38,000 (2000)	154, 174
Chukchi	ckt	Chukotko-Kamchatkan	Northern Chukotko-Kamchatkan	Russia	7,740 (2002)	114, 121, 228, 241, 267
Chumash, Barbareño (Barbareño)	boi	Chumash	Chumash	USA	extinct	276, 292
Columbia-Wenatchi	col	Salishan	Interior Salish	USA	75 (1990)	88, 90
Comanche	com	Uto-Aztecan	Numic	USA	200 (2000)	192, 221, 228, 258, 268
Coos, Hanis (Coos)	csz	Oregon Coast	Coosan	USA	1 (1964)	114, 142
Cree, Plains	crk	Algic	Algonquian	Canada	34,100 (2001)	154, 167, 228, 271
Croatian	hrv	Indo-European	Slavic	Croatia	5,546,590 (2001)	60, 79, 384, 411
Cubeo	cub	Tucanoan	Tucanoan	Colombia	6,800 (2001)	276, 286
Cuiba	cui	Guahiban	Guahiban	Colombia	2,830 (2001)	228, 258
Cuitlatec	---	Isolate	Isolate	Mexico	NA	384, 415
Cupeño	cup	Uto-Aztecan	Takic	USA	extinct	192, 200
Dagaare (Dagaare, Southern)	dga	Niger-Congo	Gur	Ghana	700,000 (2003)	154, 170–171
Dahalo	dal	Afro-Asiatic	Southern Cushitic	Kenya	400 (1992)	60, 74–75
Dakota	dak	Siouan	Siouan	USA	19,280 (1997, 2000)	60, 83

Language	ISO	Family	Genus	Location	Speakers	Page
Danish	dan	Indo-European	Germanic	Denmark	5,581,690 (2007)	14, 22, 60, 84, 114, 140, 143
Dargwa	dar	Nakh-Daghestanian	Lak-Dargwa	Russia	516,490 (2002)	306, 313
Dâw	kwa	Nadahup	Nadahup	Brazil	83 (1994)	154, 173
Diegueño, Mesa Grande (Kumiai)	dih	Hokan	Yuman	USA	330 (1991)	276, 291
Dime	dim	Afro-Asiatic	South Omotic	Ethiopia	6,197 (1994)	114, 123, 125
Dinka, Agar (dialect of Dinka, South Central)	dib	Nilo-Saharan	Nilotic	Sudan	250,000 (1956)	100, 154, 180, 276, 292
Dinka, Lanyjang (Dinka, Northeastern)	dip	Nilo-Saharan	Nilotic	Sudan	320,000 (1986)	60, 79
Diu Indo-Portuguese (Indo-Portuguese)	idb	Creole	Portuguese-lexified	India	180 (Cardoso 2013)	344, 372
Domari	rmt	Indo-European	Indic	Jordan, Iran, Israel, …	3,952,810 (2000)	384, 410
Dumi	dus	Sino-Tibetan	Bodic	Nepal	2,000 (2002)	344, 357
Dumo (Vanimo)	vam	Skou	Western Skou	Papua New Guinea	2,670 (2000)	60, 66
Dutch	nld	Indo-European	Germanic	The Netherlands	21,730,290 (2007)	xxii, 5, 14, 24–25, 29
Dutch Sign Language	dse	Sign Language	Sign Language	Netherlands	20,000 (1986)	14, 29, 60, 86, 88, 112, 190, 306, 340, 384, 415–416, 423
Dutch, Zeeuws (Zeeuws)	zea	Indo-European	Germanic	Netherlands, Belgium	220,00 (n.y.)	154, 166
Dyirbal	dbl	Australian	Pama-Nyungan	Australia	40 (1983)	60, 65, 344, 375–377
Ejagham	etu	Niger-Congo	Bantoid	Cameroon, Nigeria	116,700 (2000)	60, 76
Engenni	enn	Niger-Congo	Edoid	Nigeria	20,000 (1980)	192, 213–214, 495

Language	ISO	Family	Genus	Location	Speakers	Page
English	eng	Indo-European	Germanic	United Kingdom, Ireland, USA, Canada Australia, New Zealand, …	328,008,138 (2005)	xxii, 5–9, 14, 16–19, 23–28, 32, 38, 41, 52, 55–56, 60, 62, 67, 70–71, 74–79, 83, 88–93, 97, 99–100, 102, 106–108, 110, 114, 116–118, 121–127, 130, 133, 135–150, 154, 156–159, 162–166, 168–169, 174–176, 181–183, 187, 192, 194, 196–199, 202–205, 207, 211–214, 219–220, 222–224, 228, 234–237, 239–241, 246–249, 258, 262, 265, 267–272, 276–277, 281–283, 286, 290–294, 298–300, 303, 306, 308, 310–311, 313–314, 316–318, 320, 323–327, 329–330, 332, 334, 344, 348, 352–358, 364, 367, 369, 372–373, 375, 384–387, 391–393, 395–397, 399–400, 402–405, 410, 417–418, 423
English, American	---	Indo-European	Germanic	USA	215,000,000 (2000)	1, 7, 52, 75–76, 332, 397, 405
English, Middle	---	Indo-European	Germanic	–	extinct	384, 392
English, Old	---	Indo-European	Germanic	–	extinct	384–385, 391–392, 395–397, 410
Epena Pedee (Epena)	sja	Choco	Choco	Colombia	3,550 (2004)	228, 240, 344, 373
Ere	twp	Austronesian	Oceanic (Eastern Malayo-Polynesian)	Papua New Guinea	1,030 (1980)	75
Erromangan (Sie)	erg	Austronesian	Oceanic	Vanuatu	1,900 (2001)	88, 101–102
Eskimo Pidgin	---	Pidgin	Eskimo-lexified	Siberia, Alaska, Canada, W. Greenland	extinct (van der Voort 2013)	154, 163, 228, 249, 276
Estonian	est	Uralic	Finnic	Estonia	1,048,660 (1989)	384, 403, 412
Evenki	evn	Altaic	Tungusic	Russia	27,615 (1999)	276, 290, 306, 329
Ewe (Éwé)	ewe	Niger-Congo	Kwa	Togo, Ghana	3,112,000 (2003)	384, 394–395, 406, 408

Language	ISO	Family	Genus	Location	Speakers	Page
Fa d'Ambô (Fa D'ambu)	fab	Creole	Portuguese-lexified	Equatorial Guinea	4500–5000 (Post 2013)	154, 163
Fanakalo (Fanagalo)	fng	Pidgin	Zulu-lexified	South Africa	ca 3,000,000 (Mesthrie & Surek-Clark 2013)	114, 139, 154, 163
Faroese	fao	Indo-European	Germanic	Faroe Islands (Denmark)	48,260 (2007)	xxii, 6, 38, 45
Fijian	fij	Austronesian	Oceanic	Fiji	336,960 (1996)	88, 97, 154, 182–183, 228, 264, 276, 298–299
Finnish	fin	Uralic	Finnic	Finland	5,009,390 (1993)	114, 142, 147, 344, 349, 384, 412
Finnish Sign Language	fse	Sign Language	Sign Language	Finland	5,000 (1986)	14, 29, 33, 344, 380
Flemish Sign Language	vgt	Sign Language	Sign Language	Belgium	6,000 (2005)	14, 29
Foe (Foi)	foi	Trans-New Guinea	Kutuban	Papua New Guinea	2,800 (1980)	192, 217–218, 492
French	fra	Indo-European	Romance	France	67,838,450 (2005)	xxii, 5, 24–25, 38, 60, 67, 76, 88, 93, 114, 116, 123–124, 136, 154, 167–168, 179, 185–186, 192, 207, 211–212, 228, 249, 306, 312–313, 344, 350, 358, 370, 384, 390, 394–395, 402–403, 405, 410
French Sign Language	fsl	Sign Language	Sign Language	France	50–100,000 (1986)	14, 29
Fula, Nigerian (Fulfulde, Nigerian)	fuv	Niger-Congo	Northern Atlantic	Nigeria	1,710,000 (2000)	154, 167
Fyem (Fyam)	pym	Niger-Congo	Platoid	Nigeria	3000 (n.y.)	60, 73
Gaagudju (Gagadu)	gbu	Australian	Gaagudju	Australia	6 (1981)	344, 348
Gaelic, Scots (Gaelic, Scottish)	gla	Indo-European	Celtic	United Kingdom	66,780 (2003)	60, 77
Galician	glg	Indo-European	Romance	Spain	3,185,000 (1986)	60, 73
Gapapaiwa	pwg	Austronesian	Oceanic	Papua New Guinea	3,000 (2007)	154, 179
Ge'ez	gez	Afro-Asiatic	Semitic	(Ethiopia, Eritrea)	extinct	384, 413
Gela	nlg	Austronesian	Oceanic	Solomon Islands	11,900 (1999)	276, 290

Language	ISO	Family	Genus	Location	Speakers	Page
Georgian	kat	Kartvelian	Kartvelian	Georgia	4,255,270 (1993)	60, 78, 83
German	deu	Indo-European	Germanic	Germany	90,294,110 (1990)	8–9, 14, 24, 27, 30, 60, 74, 76, 88, 91, 106–108, 114, 123, 127, 134, 137–138, 150, 154, 165, 167, 184, 187, 192, 207, 228, 244, 266, 276, 288, 306, 317, 324, 330, 344, 370, 384, 403, 412
German Sign Language	gsg	Sign Language	Sign Language	Germany	50,000 (1986)	7, 14, 29–30, 60, 85, 88, 111, 114, 151, 154, 188–190, 192, 223–225, 276, 303–304, 384, 415–417, 422
German, Baltic (Plautdietsch)	pdt	Indo-European	Germanic	Germany	90,000 (1996)	384, 412
German, Pennsylvania	pdc	Indo-European	Germanic	USA	100,000 (2000)	384, 400, 405
Ghanaian Pidgin English	---	Pidgin	English-lexified	Ghana	ca 5,000,000 (Huber 2013)	14, 26–27, 192, 203
Ghanaian Sign Language	gse	Sign Language	Sign Language	Ghana	NA	xx, 38, 58
Gimira (Bench')	bcq	Afro-Asiatic	North Omotic	Ethiopia	174,000 (1994)	60, 81
Godié	god	Niger-Congo	Kru	Ivory Coast	26,400 (1993)	306, 323
Greek, Ancient	grc	Indo-European	Greek	–	extinct	38, 40
Greek, Homeric	grc	Indo-European	Greek	Greece	extinct	306, 313
Greek, Modern	ell	Indo-European	Greek	Greece	13,084,490 (2002)	xxii, 13–14, 17–18, 38, 40, 60, 84, 88, 94, 154, 177, 187–188, 192, 207, 211, 277, 384, 411–412
Greek Sign Language	gss	Sign Language	Sign Language	Greece	42,600 (1986)	228, 273, 344, 378
Greenlandic, West (Inuktitut, Greenlandic)	kal	Eskimo-Aleut	Eskimo		57,800 (1995)	114, 118, 154, 177, 179, 228, 268
Guaraní (Guaraní, Paraguayan)	gug	Tupian	Tupi-Guaraní	Paraguay	4,850,000 (1995)	192, 194, 384, 403, 405, 410
Gullah (Sea Island Creole English)	gul	Creole	English-lexified	USA	"no more than 10,000 monolinguals" (Klein 2013)	154, 174

Language	ISO	Family	Genus	Location	Speakers	Page
Gumuz	guk	Nilo-Saharan	Gumuz	Ethiopia, Sudan	160,000 (1994)	384, 413
Gününa Küne (Puelche)	pue	Chon	Puelche	Argentina	5 (n.y.)	276, 291
Gurindji Kriol	---	Mixed Language	Gurindji + Kriol	Australia	1000 (Meakins 2013)	228, 243
Guugu Yimidhirr (Guuguyimidjir)	kky	Australian	Pama-Nyungan	Australia	740 (1996)	344, 376–377
Hadza	hts	Isolate	Isolate	Tanzania	800 (2000)	60, 74–75
Haitian Creole (Haitian)	hat	Creole	French-lexified	Haiti	ca 11,000,000 (Fattier 2013)	60, 70, 306, 311, 384, 395
Hakka (Chinese, Hakka)	hak	Sino-Tibetan	Chinese	China	30,032,520 (1984)	60, 68, 306, 331
Halkomelem	hur	Salishan	Central Salish	Canada	225 (2002)	228, 268
Hausa	hau	Afro-Asiatic	West Chadic	Niger, Nigeria	24,988,000 (1991)	114, 129, 192, 205
Hausa Sign Language	hsl	Sign Language	Sign Language	Nigeria	NA	154, 190
Hawai'i Creole English	hwc	Creole	English-lexified	USA (Hawai'i)	600,000 (1986)	xxii, 7, 14, 26–27, 32, 88, 107, 154, 164–165, 192, 213–214, 344, 353
Hawai'i Pidgin Sign Language	hps	Sign Language	Sign Language	USA (Hawai'i)	"a few" (1987)	38, 57
Hawaiian	haw	Austronesian	Oceanic	USA (Hawai'i)	1000 (1995)	xxii, 6, 26
Hdi	xed	Afro-Asiatic	Biu-Mandara	Nigeria, Cameroon	29,000 (2001)	114, 136, 228, 245
Hebrew, Biblical	hbo	Afro-Asiatic	Semitic	–	extinct	38, 40
Hebrew, Modern	heb	Afro-Asiatic	Semitic	Israel	5,316,700 (1998)	14, 23, 60, 73, 88, 90, 98, 104–108
Hindi	hin	Indo-European	Indic	India	181,676,620 (1991)	60, 62, 70, 192, 226, 228, 241
Hixkaryana	hix	Cariban	Cariban	Brazil	600 (2000)	14, 32, 38, 51, 114, 132, 276, 286
Hong Kong Sign Language	hks	Sign Language	Sign Language	China	20,000 (2007)	14, 29, 344, 380
Hopi	hop	Uto-Aztecan	Hopi	USA	5,260 (2000)	60, 76
Hua (Yagaria, Hua)	ygr	Trans-New Guinea	Eastern Highlands	Papa New Guinea	21,100 (1982)	60, 78
Huli	hui	Trans-New Guinea	Engan	Papua New Guinea	70,000 (1991)	114, 143

Language	ISO	Family	Genus	Location	Speakers	Page
Hungarian	hun	Uralic	Ugric	Hungary	12,501,270 (2001)	154, 178–179, 228, 239, 276, 294, 306, 319
Hup (Hupdë)	jup	Nadahup	Nadahup	Colombia, Brazil	1360 (1995)	114, 173, 384, 407, 409
Iaai	iai	Austronesian	Oceanic	New Caledonia	1,560 (1996)	60, 76
Iau	tmu	Lakes Plain	Lakes Plain	Indonesia	600 (2000)	88, 99
Icelandic	isl	Indo-European	Germanic	Iceland	238,050 (1980)	60, 62, 65, 192, 207
Icen (Etykwan)	ich	Niger-Congo	Jukunoid (Benue-Congo)	Nigeria	50,200 (2000)	75
Igbo	ibo	Niger-Congo	Igboid	Nigeria	18,000,000 (1999)	114, 128, 228, 240
Ilgar (Garig-Ilgar)	ilg	Australian	Iwaidjan	Australia	extinct	114, 127
Ilocano	ilo	Austronesian	Northern Philippines	Philippines	6,996,600 (2000)	88, 92, 102
Indonesian	ind	Austronesian	Malayic	Indonesia	23,187,680 (2000)	88, 91, 154, 183, 344, 371, 384, 408–410
Indo-Pakistani Sign Language (Indian Sign Language, Pakistani Sign Language)	ins, pks	Sign Language	Sign Language	India, Pakistan	2,680,000 in India (2003); NA in Pakistan	xx, 88, 112, 114, 150–151, 192, 225–226, 228, 273, 276, 302, 306, 341, 344, 379–381, 384, 417
Ingush	inh	Nakh-Daghestanian	Nakh	Russia (Ingushetia)	413,000 (2002)	88, 94
Iquito	iqu	Zaparoan	Zaparoan	Peru	35 (2002)	228, 250
Iraqw	irk	Afro-Asiatic	Southern Cushitic	Tanzania	462,000 (2001)	154, 170
Irish (Gaelic, Irish)	gle	Indo-European	Celtic	Ireland	391,470 (1983)	60, 77, 276, 285, 384, 400
Irish Sign Language	isg	Sign Language	Sign Language	Ireland	NA	14, 29, 384, 418
Israeli Sign Language	isr	Sign Language	Sign Language	Israel	5000 (1986)	14, 29, 60, 86, 88, 110–111, 154, 190, 192, 224–225, 228, 273, 344, 378, 380, 384, 416
Italian	ita	Indo-European	Romance	Italy	61,696,677 (n.y.)	14, 22, 60, 84, 88, 93–94, 99, 114, 125, 150, 154, 185, 192, 221, 228, 247, 258–259, 276, 283, 344, 353, 366

Language	ISO	Family	Genus	Location	Speakers	Page
Italian Sign Language	ise	Sign Language	Sign Language	Italy	NA	xx, 14, 29, 154, 189–190, 306, 341, 344, 378–379, 384, 415
Jakaltek (Jacalteco, Eastern)	jac	Mayan	Mayan	Guatemala	11,000 (1998)	228, 268
Jamaican (Jamaican Creole English)	jam	Creole	English-lexified	Jamaica	3,000,000 (Farquharson 2013)	114, 141
Jaminjung (Djamindjung)	djd	Australian	Jaminjungan	Australia	30 (1991)	114, 123, 132, 384, 410
Jamsay (Dogon, Jamsay)	djm	Niger-Congo	Dogon	Mali	130,000 (1998)	154, 180
Japanese	jpn	Japanese	Japanese	Japan	122,080,100 (1985)	14, 27, 60, 62, 65, 228, 262, 335, 384, 397
Japanese Sign Language	jsl	Sign Language	Sign Language	Japan	317,000 (1986)	14, 29, 276, 304
Javanese, Peranakan	jav	Austronesian	Javanese	Indonesia	84,608,470 (2000)	114, 147
Jordanian Sign Language	jos	Sign Language	Sign Language	Jordan	NA	154, 190, 276, 303, 306, 340
Juba Arabic (Arabic, Sudanese Creole)	pga	Pidgin	Arabic-lexified	Sudan	NA (Manfredi & Petrollino 2013)	344, 372
Ju|'hoan	ktz	Khoisan	Northern Khoisan	Angola, Namibia, Botswana	33,600 (2002)	154, 168–169, 192, 196–197
Kabardian	kbd	North-West Caucasian	North-West Caucasian	Russia	1,632,500 (2002)	344, 354
Kala Lagaw Ya, Saibai	mwp	Australian	Pama-Nyungan	Australia	930 (1996)	228, 255
Kalkatungu (Kalkutung)	ktg	Australian	Pama-Nyungan	Australia	extinct	250
Kambera	xbr	Austronesian	Central Malayo-Polynesian	Indonesia	235,000 (2000)	114, 136, 154, 173–174, 228, 259
Kamula	xla	Trans-New Guinea	Kamula	Papua New Guinea	800 (1998)	306, 320
Kanoé	kxo	Kapixana	Kapixana	Brazil	No known speakers (n.y.)	228, 259
Karaim	kdr	Altaic	Turkic	Lithuania	30 (Eva Csató, p.c.)	384, 412
Karelian	krl	Uralic	Finnic	Russia	45,000 (1993)	384, 412
Karen, Pwo (Karen, Pwo Eastern)	kjp	Sino-Tibetan	Karen	Myanmar	1,050,000 (1998)	228, 259

Language	ISO	Family	Genus	Location	Speakers	Page
Karimojong (Karamojong)	kdj	Nilo-Saharan	Nilotic	Uganda	260,000 (2002)	384, 407
Karó, Arára (Karo)	arr	Tupian	Ramarama	Brazil	170 (2004)	276, 294
Kashmiri	kas	Indo-European	Indic	Pakistan, India	5,640,940 (n.y.)	154, 177, 344, 356
Kasong	---	Austro-Asiatic	Pearic	Thailand	50 (2003); NB: not the same as related but not mutually intelligible Chong [cog] (Sunee 2003)	88, 104
Kata Kolok (Benkala Sign Language)	bqy	Sign Languages	Sign Languages	Indonesia	NA	14, 30, 38, 58, 114, 151, 344, 379
Kawi (Old Javanese)	---	Austronesian	Javanese	Indonesia, Java	extinct	154, 159
Kayah Li, Eastern (Kayah, Eastern)	eky	Sino-Tibetan	Karen	Myanmar, Thailand	26,000 (2007)	60, 81, 306, 327
Kayardild	gyd	Australian	Tangkic	Australia	150 (2000)	60, 83, 114, 143, 192, 213, 228, 245–246, 365–366
Kefa (Kafa)	kbr	Afro-Asiatic	South Omotic	Ethiopia	570,000 (1994)	384, 413
Kele	sbc	Austronesian	Oceanic (Eastern Malayo-Polynesian)	Papua New Guinea	600 (1982)	75
Kenyan Sign Language	xki	Sign Language	Sign Language	Kenya	NA	14, 29
Kham (Kham, Gamale)	kgj	Sino-Tibetan	Bodic	Nepal	13,100 (2000)	228, 245, 256
Khanty	kca	Uralic	Ugric	Russia	13,600 (n.y.)	60, 79
Khmer (Khmer, Central)	khm	Austro-Asiatic	Khmer	Cambodia	13,603,400 (2006)	114, 140
Khwarshi (Khvarshi)	khu	Nakh-Daghestanian	Avar-Andic-Tsezic	Russia	1,870 (n.y.)	154, 186, 276, 295–297, 306, 310
Kikongo-Kituba (Kituba)	ktu	Creole	Kikongo-Kimanyanga-lexified	DR Congo	6–8.000.000 (Mufwene 2013 & pc)	154, 171
Kiowa	kio	Kiowa Tanoan	Kiowa Tanoan	USA	1,100 (2000)	228, 250
Kipea (Karirí-Xocó)	kzw	Kariri	Kariri	Brazil	No known speakers (1995)	88, 97

Language	ISO	Family	Genus	Location	Speakers	Page
Kisi (Kissi, Northern)	kqs	Niger-Congo	Southern Atlantic	Guinea	327,000 (1991)	88, 99, 108, 154, 167, 228, 250
Klao	klu	Niger-Congo	Kru	Liberia	213,000 (2006)	60, 66–67, 384, 396
Koasati	cku	Muskogean	Muskogean	USA	200 (2000)	114, 137
Kobon	kpw	Trans-New Guinea	Madang	Papua New Guinea	10,000 (2007)	114, 143
Kom	bkm	Niger-Congo	Bantoid (Benue-Congo)	Cameroon	233,000 (2005)	75
Kombai	tyn	Trans-New Guinea	Awju-Dumut	Indonesia	4000 (1991)	306, 329
Kongo (Koongo)	kng	Niger-Congo	Bantoid	DR Congo	5,000,000	192, 200
Konkani	knn	Indo-European	Indic	India	4,004,490 (n.y.)	14, 33
Kono	kno	Niger-Congo	Western Mande	Sierra Leone	205,000 (2006)	384, 394
Konzime (Koonzime)	ozm	Niger-Congo	Bantoid	Cameroon	30,000 (2000)	276, 292
Korean	kor	Isolate	Isolate	N, S Korea	66,305,890 (1986)	14, 22, 114, 144, 344, 354, 372–374
Korlai (Korlai Creole Portuguese)	vkp	Creole	Portuguese-lexified	India	750 (1998)	114, 141, 154, 163, 344, 372
Koro	kxr	Austronesian	Oceanic	Papua New Guinea	400 (1983)	75
Koromfe	kfz	Niger-Congo	Gur	Burkina Faso, Mali	198,000 (2001)	114, 136, 228, 241–242
Koyra Chiini (Songhay, Koyra Chiini)	khq	Nilo-Saharan	Songhay	Mali	200,000 (1999)	60, 67, 88, 97
Koyraboro Senni (Songhay, Koyraboro Senni)	ses	Nilo-Saharan	Songhai	Mali	300,000 (2007)	228, 268
Kriol	rop	Creole	English-lexified	Australia	10,000 (1991)	228, 243, 384, 410
Krongo	kgo	Kadugli	Kadugli	Sudan	21,700 (1984)	114, 142, 228, 268
Kuche (Che)	ruk	Niger-Congo	Platoid	Nigeria	100,000 (2003)	192, 199
Kunjen	kjn	Australian	Pama-Nyungan	Australia	20 (1991)	60, 77
Kuot	kto	Isolate	Isolate	Papua New Guinea	2,400 (2002)	228, 259
Kupto (Kutto)	kpa	Afro-Asiatic	West Chadic	Nigeria	3,000 (1995)	384, 393
Kurti	ktm	Austronesian	Oceanic	Papua New Guinea	3,000 (2002)	75
Kutenai	kut	Isolate	Isolate	Canada, USA	12 (2002)	114, 142, 276, 291

Language	ISO	Family	Genus	Location	Speakers	Page
Kuuk Thaayorre (Thayore)	thd	Australian	Pama-Nyungan	Australia	150 (1991)	114, 134
Lak	lbe	Nakh-Daghestanian	Lak-Dargwa	Russia	164,420 (2002)	154, 168
Lakhota (Lakota)	lkt	Siouan	Siouan	USA	6,390 (1997)	114, 144
Lango	laj	Nilo-Saharan	Nilotic	Uganda	1,490,000 (2002)	88, 99–100, 114, 117–118, 130, 344, 365
Lardil	lbz	Australian	Pama-Nyungan	Australia	2 (2000)	60, 66
Larike (Larike-Wakasihu)	alo	Austronesian	Central Malayo-Polynesian	Indonesia	12,600 (1987)	154, 161
Latin	lat	Indo-European	Italic	–	extinct	5–6, 88, 96–97, 306, 313, 344, 370, 384, 394, 398, 408
Latvian	lav	Indo-European	Baltic	Latvia	1,504,880 (n.y.)	60, 80, 192, 221–222, 384, 412
Lavukaleve	lvk	Solomons East Papuan	Lavukaleve	Solomons Islands	1,780 (1999)	38, 55, 114, 123, 154, 160, 344, 358, 362
Lebanese Sign Language	---	Sign Language	Sign Language	Lebanon	NA	344, 378–379
Leipon	lek	Austronesian	Oceanic	Papua New Guinea	650 (1977)	75
Lele	lle	Austronesian	Oceanic	Papua New Guinea	1,300 (1982)	75
Leti	lti	Austronesian	Central Malayo-Polynesian	Indonesia	7,500 (1995)	88, 91–92
Lezgian (Lezgi)	lez	Nakh-Daghestanian	Lezgic	Russia	783,720 (2002)	154, 177–179, 306, 311–312, 344, 365, 384, 395
Lihir	lih	Austronesian	Oceanic	Papua New Guinea	12,600 (2000)	154, 161–162
Likpe (Sekpele)	lip	Niger-Congo	Kwa	Ghana	23,400 (2003)	384, 406, 408–409
Limbu	lif	Sino-Tibetan	Bodic	Bhutan, India, Nepal	421,500 (2001)	114, 145
Lingala	lin	Creole	Bobangi-lexified	DR Congo	ca 15,000,000 (Meeuwis 2013)	154, 171, 228, 249
Lithuanian	lit	Indo-European	Baltic	Lithuania	3,154,180 (1998)	384, 412
Liv	liv	Uralic	Finnic	Latvia	15 (1995)	384, 412

Language	ISO	Family	Genus	Location	Speakers	Page
Logba	lgq	Niger-Congo	Kwa	Ghana	7,500 (2003)	88, 91, 306, 333
Logoti (Logo)	log	Nilo-Saharan	Moru-Ma'di	DR Congo	210,000 (1989)	306, 336
Lokaa	yaz	Niger-Congo	Cross River	Niger	120, 000 (1989)	306, 334, 344, 350
Loniu	los	Austronesian	Oceanic	Papua New Guinea	460 (1977)	306, 333
Lugbara	lgg	Nilo-Saharan	Moru-Ma'di	DR Congo, Uganda	1,637,000 (2004)	88, 98
Luo (Dholuo)	luo	Nilo-Saharan	Nilotic	Kenya, Tanzania	4,410,000 (n.y.)	384, 409
Ma'a/Mbugu (Mbugu)	mhd	Mixed Language	Bantu + Cushitic	Tanzania	7000 (2003)	228, 249, 344, 366–367
Maasai	mas	Nilo-Saharan	Nilotic	Kenya, Tanzania	1,045,000 (n.y.)	88, 98, 114, 122, 154, 180
Maba	mde	Nilo-Saharan	Maban	Chad	296,000 (2006)	154, 180
Macedonian	mkd	Indo-European	Slavic	Macedonia	2,113,170 (1986)	384, 411
Macushi	mbc	Cariban	Cariban	Brazil, Guyana, Venezuela	29,100 (2001)	114, 147, 276, 289, 344, 365
Maithili	mai	Indo-European	Indic	India, Nepal	34,700,000 (2000)	344, 364
Makaa	mcp	Niger-Congo	Bantoid	Cameroon	80,000 (1987)	344, 350
Malagasy (Malagasy, Plateau)	plt	Austronesian	Barito	Madagascar	7,528,900 (2006)	137, 276, 293
Malakmalak (Mullukmulluk)	mpb	Australian	Northern Daly	Australia	1920 (2001)	60, 76
Malay	zlm	Austronesian	Malayic	Malaysia	10,296,000 (n.y.)	14, 24, 384, 408
Malayalam	mal	Dravidian	Southern Dravidian	India	35,893,990 (1997)	60, 83, 344, 365
Malto (Kumarbhag Paharia)	kmj	Dravidian	Northern Dravidian	India	20,200 (2000)	114, 118
Mamaindê	wmd	Nambikuaran	Nambikuaran	Brazil	330 (2007)	88, 110, 113–114, 122, 135
Manam	mva	Austronesian	Oceanic	Papua New Guinea	7,950 (2003)	154, 161
Manange (Manangba)	nmm	Sino-Tibetan	Bodic	Nepal	3740 (1988)	384, 406
Mandarin (Chinese, Mandarin)	cmn	Sino-Tibetan	Chinese	China	845,456,760 (2000, 2005)	14, 16–19, 28, 60, 76, 88, 96, 331, 344, 355, 384, 398
Mandinka	mnk	Niger-Congo	Western Mande	Mali, Senegal, Guinea	1,346,000 (2006)	384, 395
Mangarrayi (Mangarayi)	mpc	Australian	Mangarrayi	Australia	50 (1983)	276, 286

Language	ISO	Family	Genus	Location	Speakers	Page
Mangbetu	mdj	Nilo-Saharan	Mangbetu	DRC	620,000 (n.y.)	75
Maninka (Maninkakan, Eastern)	emk	Niger-Congo	Western Mande	Guinea	2,031,800 (1986)	384, 396
Mantjiltjara (Martu Wangka, Mantjiltjara)	mpj	Australian	Pama-Nyungan	Australia	720 (1991)	60, 78
Maori	mri	Austronesian	Oceanic	New Zealand	60,260 (1995)	14, 23, 114, 118
Mapudungun	arn	Araucanian	Araucanian	Chile	300,039 (2002)	192, 216
Maranao	mrw	Austronesian	Southern Philippines	Philippines	776,000 (1990)	88, 91
Marathi	mar	Indo-European	Indic	India	68,061,130 (1997)	228, 241
Margi	mrt	Afro-Asiatic	Biu-Mandara	Nigeria	158,000 (2006)	60, 72
Maricopa	mrc	Yuman	Yuman	USA	160 (2000)	114, 137, 228, 240
Maritime Sign Language	nsr	Sign Language	Sign Language	Canada	NA	38, 57
Martuthunira (Martuyhunira)	vma	Australian	Pama-Nyungan	Australia	5 (1981)	306, 329
Matsés	mcf	Panoan	Panoan	Brazil, Peru	2,200 (2006)	192, 218–219
Matuumbi (Matumbi)	mgw	Niger-Congo	Bantoid	Tanzania	72,000 (1978)	276, 285
Mauwake	mhl	Trans-New Guinea	Madang	Papua New Guinea	2,390 (2003)	228, 259, 306, 322, 335–336, 343
Maybrat (Mai Brat)	ayz	West Papuan	North-Central Bird's Head	Indonesia	20,000 (1987)	88, 106, 192, 198
Mbabaram	vmb	Australian	Pama-Nyungan	Australia	extinct (2003)	60, 65
Mbay	myb	Nilo-Saharan	Bongo-Bagirmi	Chad	88,300 (1990)	114, 135
Mbili (Bambili-Bambui)	baw	Niger-Congo	Bantoid	Cameroon	10,000 (1983)	276, 287
Media Lengua	mue	Mixed Language	Spanish + Quechua	Ecuador	Not known but more than 200 (Muysken 2013)	14, 27, 114, 141, 228, 249
Mehek	nux	Sepik	Tama Sepik	Papua New Guinea)	6,300 (1994)	154, 156
Meithei (Meitei)	mni	Sino-Tibetan	Kuki-Chin	India	1,391,000 (2000)	88, 104–105
Mexican Sign Language	mfs	Sign Language	Sign Language	Mexico	87,000–100,000 (1986)	192, 224

Language	ISO	Family	Genus	Location	Speakers	Page
Michif	crg	Mixed Language	French + Plains Cree	USA	200–2,000 (Bakker 2013)	14, 26, 154, 171, 228, 249, 276, 286
Minangkabau	min	Austronesian	Malayic	Indonesia	5,530,000 (2007)	154, 172
Mixtec, Alacatlatzala	mim	Oto-Manguean	Mixtecan	Mexico	22,200 (2000)	306, 321
Mixtec, Chalcatongo (Mixtec, San Miguel el Grande)	mig	Oto-Manguean	Mixtecan	Mexico	14,453 (1990)	344, 356
Mocoví	moc	Guaicuruan	Guaicuruan	Argentina	4,530 (2000)	114, 140
Mopun (Mwaghavul, Mupun)	sur	Afro-Asiatic	West Chadic	Nigeria	295,000 (1993)	384, 394
Mosetén (Tsimané)	cas	Mosetenan	Mosetenan	Bolivia	585 (2000)	384, 406
Mualang	mtd	Austronesian	Malayic	Indonesia	40,000 (2007)	228, 265–266, 306, 318
Mugil (Bargam)	mlp	Trans-New Guinea	Madang	Papua New Guinea	3,750 (2003)	276, 297–298
Muna	mnb	Austronesian	Sulawesi	Indonesia	300,000 (1989)	75
Murle	mur	Nilo-Saharan	Surmic	Sudan	60,200 (1982)	228, 239–240
Mwotlap (Motlav)	mlv	Austronesian	Oceanic	Vanuatu	2,200 (2001)	114, 125–126
Nadëb	mbj	Nadahup	Nadahup	Brazil	300 (1986)	114, 120, 228, 269, 276, 286
Naga Pidgin	nag	Creole	Assamese-lexified	India	30,000 (1989)	344, 366
Nahuatl, Huasteca (Nahuatl, Eastern Huasteca)	nhe	Uto-Aztecan	Aztecan	Mexico	410,000 (1991)	114, 120
Nakanai	nak	Austronesian	Oceanic	Papua New Guinea	13,000 (1981)	88, 103
Nali	nns	Austronesian	Oceanic	Papua New Guinea	1,800 (1982)	75
Nalik	nal	Austronesian	Oceanic	Papua New Guinea	5,140 (1990)	154, 157, 192, 212
Nandi	niq	Nilo-Saharan	Nilotic	Kenya	NA	88, 98, 154, 180
Natügu	stc	Austronesian	Oceanic	Solomon Islands	4,280 (1999)	60, 76
Navajo	nav	Na-Dene	Athapaskan	USA	149,000 (1990)	38, 45, 114, 137, 276, 300–301
Naʔahai (South West Bay)	sns	Austronesian	Oceanic	Vanuatu	600 (2001)	75

Language	ISO	Family	Genus	Location	Speakers	Page
Ndom	nqm	Kolopom	Kolopom	Indonesia	1,200 (2002)	114, 143
Ndyuka (Aukan)	djk	Creole	English-lexified	Suriname	60,500 (Migge 2013)	114, 142, 145, 154, 171, 174, 192, 228, 248, 266, 276, 286, 291, 293–294, 306, 329, 344, 366
Nepali	nep	Indo-European	Indic	Nepal	13,875,700 (2001)	344, 371, 384, 406
New Zealand Sign Language	nzs	Sign Language	Sign Language	Nez Zealand	NA	14, 29, 114, 152
Nez Perce	nez	Penutian	Sahaptian	USA	200 (1997)	228, 241, 268
Ngemba	nge	Niger-Congo	Bantoid	Cameroon	18,800 (2002)	306, 327
Ngiti	niy	Nilo-Saharan	Lendu	DR Congo	100,000 (1991)	114, 143
Ngiyambaa (Wangaaybuwan-Ngiyambaa)	wyb	Australian	Pama-Nyungan	Australia	12 (1981)	154, 177
Ngwe	nwe	Niger-Congo	Bantoid	Cameroon	73,200 (2001)	75
Nhanda	nha	Australian	Pama-Nyungan	Australia	NA	276, 282
Nias	nia	Austronesian	Sundic	Indonesia	770,000 (2000)	75
Nicobarese, Car	caq	Austro-Asiatic	Nicobarese	India	37,000 (2005)	228, 258
Nigerian Pidgin English	pcm	Pidgin	English-lexified	Nigeria	ca 75,000,000 (Faraclas 2013)	14, 25
Nishi (Nisi)	dap	Sino-Tibetan	Tani	India	261,000 (1997)	60, 77
Nivkh (Gilyak)	niv	Isolate	Isolate	Russia	690 (1995)	228, 259
Nkore-Kiga (Chiga)	cgg	Niger-Congo	Bantoid	Uganda	1,580,000 (2002)	114, 149, 154, 168, 344, 355
Nocte (Naga, Nocte)	njb	Sino-Tibetan	Northern Naga	India	35,000 (2001)	228, 270
Norwegian	nor	Indo-European	Germanic	Norway	4,640,000 (n.y.)	14, 22, 60, 80, 114, 140
Nuer	nus	Nilo-Saharan	Nilotic	Sudan	804,900 (1982)	88, 100, 154, 180
Nunggubuyu	nuy	Australian	Nunggubuyu	Australia	360 (1996)	38, 51
Nuuchahnulth (Nootka)	noo	Wakashan	Southern Wakashan	Canada	200 (2002)	114, 125–126, 150
O'odham (Tohono O'odham)	ood	Uto-Aztecan	Tepiman	Mexico, USA	9,600 (2000)	88, 100

Language	ISO	Family	Genus	Location	Speakers	Page
Ojibwe, Minnesota (Chippewa)	ciw	Algic	Algonquian	USA	5000 (1990)	114, 135, 276, 292
Old Bangkok Sign Language	---	Sign Language	Sign Language	Thailand	NA	38, 58
Old Chiangmai Sign Language (Chiangmai Sign Language)	csd	Sign Language	Sign Language	Thailand	NA	38, 58
Old French Sign Language	---	Sign Language	Sign Language	France	extinct	384, 416
Old Norse	---	Indo-European	Germanic	Norway, Sweden, Denmark	extinct	114, 121, 384, 399, 410
Orokaiva	okv	Trans-New Guinea	Binanderean	Papua New Guinea	35,000 (2000)	344, 349
Oromo, Harar (Oromo, Eastern)	hae	Afro-Asiatic	Eastern Cushitic	Ethiopia	4,530,000 (1994)	192, 195, 230, 240
Otomí, Mezquital	ote	Oto-Manguean	Otomian	Mexico	100,100 (1990)	384, 407, 410
Otomí, Sierra (Otomi, Eastern Highland)	otm	Oto-Manguean	Otomian	Mexico	49,300 (2007)	88, 107–108
Pacoh	pac	Austro-Asiatic	Katuic	Vietnam	29,200 (2002)	114, 117
Palauan	pau	Austronesian	Palauan	Palau	17,140 (2000)	228, 249, 344, 353
Papiamentu	pap	Creole	Spanish -lexified	Netherlands Antilles	320,200 (1998)	114, 141, 228, 249, 306, 311, 344, 372
Papitalai	pat	Austronesian	Oceanic	Papua New Guinea	520 (1977)	75
Päri	lkr	Nilo-Saharan	Nilotic	Sudan	28,000 (1987)	228, 268, 276, 285–286
Pashai (Pashayi, Southeast)	psi	Indo-European	Iranian	Afghanistan	54,400 (2000)	228, 242
Pashto (Pashto, Central)	pst	Indo-European	Iranian	Afghanistan, Pakistan	7,920,000 (n.y.)	88, 94–95
Persian (Farsi, Western)	pes	Indo-European	Iranian	Iran	23,879,300 (1997)	xxii, 14, 23, 33, 88, 102, 154, 173
Persian, Old	---	Indo-European	Iranian	-	extinct	228, 243
Pichi (Fernando Po Creole English)	fpe	Creole	English-lexified	Equatorial Guinea	ca 100,000 (Yakpo 2013)	88, 106–108, 344, 367
Pidgin Hindustani	---	Pidgin	Fiji Hindi-lexified	Fiji	NA (Siegel 20123)	114, 141, 154, 163, 276, 286, 306
Pipil	ppl	Uto-Aztecan	Aztecan	El Salvador	20 (1987)	228, 251
Pirahã	myp	Mura	Mura	Brazil	360 (2000)	60, 71–72, 75, 114, 143, 154, 159–160, 384, 410

Language	ISO	Family	Genus	Location	Speakers	Page
Pitjantjatjara	pjt	Australian	Pama-Nyungan	Australia	2,120 (1996)	228, 238, 240, 306, 312
Plains-Indian Sign Language	psd	Sign Language	Sign Language	USA	NA	38, 58
Plang (Blang)	blr	Austro-Asiatic	Palaung-Khmuic	China	55,200 (2000)	114, 134
Pomo, Eastern	peb	Hokan	Pomoan	USA	extinct	228, 254–255, 306, 338
Ponam	ncc	Austronesian	Oceanic	Papua New Guinea	420 (1977)	75
Portuguese	por	Indo-European	Romance	Portugal, Brazil	177,981,570 (n.y.)	xxii, 5–7, 24, 38, 53, 82, 88, 94, 141
Portuguese, Brazilian	por	Indo-European	Romance	Brazil	177,981,570 (n.y.)	xxii, 6–7, 60, 94, 385, 394
Principense	pre	Creole	Portuguese-lexified	DR São Tomé & Príncipe	fewer than 100 (Maurer 2013c)	344, 372
Proto-Chadic	---	Afro-Asiatic	Chadic	S. Chad	extinct	384, 394
Proto-Germanic	---	Indo-European	Germanic	N. Germany	extinct	384, 392, 400
Proto-Indo-European	---	Indo-European	–	Russian Caucasus	extinct	384, 393
Proto-Norse	---	Indo-European	Germanic	Scandinavia	extinct	384, 398–399
Proto-Oceanic	---	Austronesian	Oceanic	Papua New Guinean Archipelago	extinct	384, 399
Providencia Sign Language	prz	Sign Language	Sign Language	Colombia	19 (1986)	38, 57
Puquina	---	Isolate	Isolate	Bolivia, Peru	extinct	344, 354
Purépecha	tsz	Tarascan	Tarascan	Mexico	40,850 (2005)	344, 364, 384, 408–411
Qafar (Afar)	aar	Afro-Asiatic	Eastern Cushitic	Ethiopia, Eritrea, Djibouti	1,078,200 (1994)	384, 413
Qiang (Qiang, North; Qiang, South)	cng, qxs	Sino-Tibetan	Quiangic	China	57,800 (1999) + 81,300 (1999)	192, 217
Quapaw	qua	Siouan	Siouan	USA	34 (1990)	60, 72–73
Quebec Sign Language	fcs	Sign Language	Sign Language	Canada	NA	14, 29, 114, 151
Quechua, Imbabura (Quichua, Imbabura Highland)	qvi	Quechuan	Quechuan	Ecuador	300,000 (1977)	14, 27, 384, 408
Quiché (K'iche', Central)	quc	Mayan	Mayan	Guatemala	1,900,000 (2000)	384, 415

Language	ISO	Family	Genus	Location	Speakers	Page
Rama	rma	Chibchan	Rama	Nicaragua	24 (1989)	60, 84
Rapanui (Rapa Nui)	rap	Austronesian	Oceanic	Easter Island, Chile	3,290 (2000)	60, 83, 88, 97, 192, 198, 306, 308–309, 384, 409
Rendille	rel	Afro-Asiatic	Eastern Cushitic	Kenya	34,700 (2006)	192, 210
Rennellese Sign Language	rsi	Sign Language	Sign Language	Solomon Islands	NA	38, 58
Réunion Creole	rcf	Creole	French-lexified	Réunion	800,000 (Bollée 2013)	228, 249, 276
Romani, Bugurdži (Romani, Balkan)	rmn	Indo-European	Indic	Serbia, Montenegro	709,570 (n.y.)	384, 411
Romanian	ron	Indo-European	Romance	Romania	23,351,080 (2002)	384, 411
Romansch	roh	Indo-European	Romance	Switzerland	35,100 (2000)	38, 53
Rotokas	roo	West Bougainville	West Bougainville	Papua New Guinea	4,320 (1981)	14, 31, 60, 63, 71–72, 74
Roviana	rug	Austronesian	Oceanic	Solomon Islands	9870 (1999)	276, 287
Rumungro (Romani, Carpathian)	rmc	Indo-European	Indic	Czech Republic	472,470 (2001)	384, 408
Rushan (Shughni, Rushani)	sgh	Indo-European	Iranian	Tajikistan	60,000 (1990)	228, 242–243
Russian	rus	Indo-European	Slavic	Russia	143,553,950 (2002)	xxii, 6, 114, 118, 140, 154, 188, 276, 294
Russian Sign Language	rsl	Sign Language	Sign Language	Russia	NA	14, 29
Sa'a	apb	Austronesian	Oceanic	Solomon Islands	11,500 (1999)	384, 399
Saami, Inari	smn	Uralic	Sami	Finland	300 (2001)	60, 79
Salinan	sln	Salinan	Salinan	USA	NA	276, 287
Sandawe	sad	Khoisan	Sandawe	Tanzania	40,000 (2000)	60, 74–75
Sango	sag	Creole	Ngbandi-lexified	Central African Republic	2,500,000 (Samarin 2013)	60, 82, 88, 106, 114, 140, 142, 145, 154, 171, 192, 228, 248–249, 266, 276, 286, 291, 293–294, 306, 311, 329, 344, 366
Sanskrit	san	Indo-European	Indic	India	extinct	38, 40, 117–118, 384, 409–410

Language	ISO	Family	Genus	Location	Speakers	Page
Sanuma	xsu	Yanomam	Yanomam	Venezuela	6,410 (2000)	228, 268
Saramaccan	srm	Creole	English-lexified	Suriname	26,000 (1995)	114, 141
Sawu (Sabu)	hvn	Austronesian	Central Malayo-Polynesian	Indonesia	110,000 (1997)	344, 374
Selknam (Ona)	ona	Chon	Chon Proper	Argentina	2 (1991)	14, 32, 276, 286
Semelai	sza	Austro-Asiatic	Aslian	Malaysia	6,418 (2003)	228, 241
Serbian	srp	Indo-European	Slavic	Serbia	7,020,550 (2006)	384, 411
Seri	sei	Hokan	Seri	Mexico	900 (2007)	60, 83, 276, 291
Sheko	she	Afro-Asiatic	North Omotic	Ethiopia	40,000 (2007)	344, 346–347
Shilluk	shk	Nilo-Saharan	Nilotic	Sudan	175,000 (1982)	154, 180
Shuswap	shs	Salishan	Interior Salish	Canada	500 (2002)	154, 183
Sidaama (Sidamo)	sid	Afro-Asiatic	East Cushitic	Ethiopia	2,900,000 (2005)	384, 413
Singapore Bazaar Malay	---	Pidgin	Malay-lexified	Singapore	fewer than 10,000 (Aye 2013)	114, 140, 228, 249, 306
Singlish	---	Creole	English-lexified	Singapore	NA	60, 82, 276, 286
Slovene	slv	Indo-European	Slavic	Slovenia	1,909,050 (1991)	60, 73, 135
Som	smc	Trans-New Guinea	Finisterre-Huon	Papua New Guinea	80 (2000)	114, 143
Somali	som	Afro-Asiatic	Eastern Cushitic	Somalia	13,871,700 (2006)	114, 148, 384, 413
Sonsorol-Tobi (Sonsorol)	sov	Austronesian	Oceanic	Palau	660 (1981)	306, 314
South Korean Sign Language (Korean Sign Language)	kvk	Sign Language	Sign Language	South Korea	NA	14, 29, 344, 380
Spanish	spa	Indo-European	Romance	Spain	328,518,810 (1986)	14, 24–25, 27, 60, 62, 78, 88, 90, 104–105, 192, 207, 215, 276, 303, 306, 309, 384, 399, 403–407, 409–411
Spanish Sign Language	ssp	Sign Language	Sign Language	Spain	102,000 (1994)	114, 152, 344, 379

Language	ISO	Family	Genus	Location	Speakers	Page
Spanish, Middle	---	Indo-European	Romance	Spain	extinct	384
Spanish, Old	---	Indo-European	Romance	Spain	extinct	384, 398–399
Sranan	srn	Creole	English-lexified	Suriname	126,400 (1993)	14, 26, 60, 82
Sri Lankan Malay (Sri Lankan Creole Malay)	sci	Creole	Malay-lexified	Sri Lanka	30–40,000 (Slomanson 2013)	114, 140, 306
Sri Lanka Portuguese (Indo-Portuguese)	idb	Creole	Portuguese-lexified	Sri Lanka	4,970 (2006)	154, 171, 174, 306
Sumerian	---	Isolate	Isolate	Iraq	extinct	38, 40
Supyire (Senoufo, Supyire)	spp	Niger-Congo	Gur	Mali	350,000 (2007)	114, 143, 192, 199, 306, 326, 344, 354
Sursurunga	sgz	Austronesian	Oceanic	Papua New Guinea	3,000 (1991)	154, 162
Swahili	swh	Niger-Congo	Bantoid	Tanzania	787,630 (2006)	114, 150, 384, 395, 402, 410
Swahili, Katanga (Swahili, Congo)	swc	Niger-Congo	Bantoid	DR Congo	1000 (n.y.)	384, 402, 410
Swedish	swe	Indo-European	Germanic	Sweden	8,311,739 (1998)	xxii, 6, 14, 22, 38, 53, 60, 76, 80, 114, 123, 131, 140, 144–145, 150, 154, 168, 192, 202, 207, 222, 228, 275–276, 306, 330, 344, 353, 370–371, 384, 389, 393, 399, 404, 412
Swedish Sign Language	swl	Sign Language	Sign Language	Sweden	8,000 (1986)	14, 29, 154, 189, 192, 224–225, 276, 296, 299, 303, 306, 339, 344, 378
Swedish, Old	---	Indo-European	Germanic	Sweden	extinct	384, 389
Tabassaran	tab	Nakh-Daghestanian	Lezgic	Russia	128,900 (2002)	154, 178
Tagalog	tgl	Austronesian	Meso-Philippine	Philippines	23,853,200 (2000)	88, 103, 114, 140
Taiwanese (Chinese, Min Nan)	nan	Sino-Tibetan	Chinese	Taiwan	47,265,100 (1988)	384, 397
Taiwanese Sign Language	tss	Sign Language	Sign Language	Taiwan	NA	14, 29, 344, 380
Tajik (Tajiki)	tgk	Indo-European	Iranian	Tajikistan	4,457,500 (1991)	384, 409
Takia	tbc	Austronesian	Oceanic	Papua New Guinea	40,000 (n.y.)	114, 144
Tariana	tae	Arawakan	Arawakan	Brazil	430 (1996)	344, 363, 384, 405, 409

Language	ISO	Family	Genus	Location	Speakers	Page
Tashlhiyt (Tachelhit)	shi	Afro-Asiatic	Berber	Morocco	3,000,000 (1998)	60, 68, 344, 360
Tauya	tya	Trans-New Guinea	Madang	Papua New Guinea	350 (1981)	344, 347
Tayo	cks	Creole	French-lexified	New Caledonia	ca 3,000 (Ehrhart & Halpap 2013)	14, 26, 276
Tepehua (Tepehua, Pisaflores)	tpp	Totonacan	Totonacan	Mexico	4,000 (1990)	276, 288
Teso	teo	Nilo-Saharan	Nilotic	Uganda	1,849,000 (2002)	384, 393–394
Thai Sign Language	tsq	Sign Language	Sign Language	Thailand	51,000 (1997)	14, 29, 38, 58, 344, 380
Thao	ssf	Austronesian	Paiwanic	Taiwan	6 (2000)	88, 102
Tibetan, Classical	---	Sino-Tibetan	Bodic	Tibet	extinct	306, 313–314
Ticuna	tca	Isolate	Isolate	Brazil, Peru	48,600 (1998)	60, 81
Tigre	tig	Afro-Asiatic	Semitic	Eritrea	8,000 (1982)	384, 413
Tigrinya (Tigrigna)	tir	Afro-Asiatic	Semitic	Ethiopia, Eritrea	4,449,875 (2001)	384, 413
Tiipay, Jamul (Kumiai, Tiipay)	dih	Hokan	Yuman	Mexico, USA	330 (1991)	114, 148–149, 276, 291, 306, 319, 326
Tinrin (Tiri)	cir	Austronesian	Oceanic	New Caledonia	260 (1996)	344, 349
Tiriyo (Trió)	tri	Cariban	Cariban	Suriname	2,300 (2003)	228, 250, 276, 286, 344, 363
Titan	ttv	Austronesian	Oceanic	Papua New Guinea	3,850 (1992)	75
Tiwi	tiw	Australian	Tiwian	Australia	1,830 (1996)	88, 106
Tlingit	tli	Na-Dene	Tlingit	USA	1,430 (2000)	228, 258
Tobati	tti	Austronesian	Oceanic	Indonesia	350 (1998)	276, 286
Tocharian A	---	Indo-European	Tocharian	W. China	extinct	154, 160
Tocharian B	---	Indo-European	Tocharian	W. China	extinct	154, 160
Tok Pisin	tpi	Pidgin	English-lexified	Papua New Guinea	3–5,000,000 (Smith & Siegel 2013)	14, 24–25
Tongan	ton	Austronesian	Oceanic	Tonga	126,390 (1998)	60, 70

Language	ISO	Family	Genus	Location	Speakers	Page
Toqabaqita (To'abaita)	mlu	Austronesian	Oceanic	Solomon Islands	12,600 (1999)	114, 129, 306, 334, 344, 358–359, 384, 395
Tsou	tsu	Austronesian	Tsouic	Taiwan	2,130 (2002)	114, 146
Tucano	tuo	Tucanoan	Tucanoan	Colombia, Brazil	4,630 (1986)	344, 362–363, 384
Tugun	tzn	Austronesian	Central Malayo-Polynesian	Indonesia	1,200 (1990)	192, 214–215
Tukang Besi (Tukang Besi, North)	khc	Austronesian	Sulawesi	Indonesia	180,000 (1995)	228, 263–264
Turkish	tur	Altaic	Turkic	Turkey	50,750,120 (1987)	38, 55, 60, 77, 88, 96, 107–111, 154, 177, 182–183, 228, 261, 306, 312, 325, 329, 344, 356, 384, 395
Turkish Sign Language	tsm	Sign Language	Sign Language	Turkey	NA	384, 408, 411
Turkish, Macedonian (Turkish, Balkan Gagauz)	bgx	Altaic	Turkic	Macedonia	4000 (n.y.)	88, 111, 344, 378, 381
Twi (Akan, Twi)	aka	Niger-Congo	Kwa	Ghana	8,300,000 (2004)	384, 397
Tzutujil (Tz'utujil, Eastern)	tzj	Mayan	Mayan	Guatemala	50,000 (1998)	154, 156–157, 184, 228, 252, 268, 344, 354–355
Udi	udi	Nakh-Daghestanian	Lezgic	Azerbaijan	8,440 (1995)	88, 94–95
Udihe	ude	Altaic	Tungusic	Russia	230 (n.y.)	154, 164, 177, 179
Ugandan Sign Language	ugn	Sign Language	Sign Language	Uganda	NA	14, 29
Ungarinjin (Ngarinyin)	ung	Australian	Wororan	Australia	82 (1981)	154, 179, 276, 286
Urarina	ura	Isolate	Isolate	Peru	3,000 (2002)	114, 149, 276, 286
Urdu	urd	Indo-European	Indic	Pakistan	60,586,800 (1993)	60, 70, 192, 226
Uripiv	upv	Austronesian	Oceanic	Vanuatu	3,450 (2001)	75
Urubú Sign Language (Kaapor Sign Language)	uks	Sign Language	Sign Language	Brazil	7 (1986)	14, 30, 276, 304
Ùsàk Èdèt (Usaghade)	usk	Niger-Congo	Cross River	Cameroon	10,000 (1990)	384, 394
Uzbek (Uzbek, Northern)	uzn	Altaic	Turkic	Uzbekistan	18,817,600 (1995)	384, 409

Language	ISO	Family	Genus	Location	Speakers	Page
Vietnamese	vie	Austro-Asiatic	Viet-Muong	Vietnam	68,634,000 (1999)	60, 77, 114, 147
Wambaya	wmb	Australian	West Barkly	Australia	12 (1981)	228, 264
Wan	wan	Niger-Congo	Eastern Mande	Ivory Coast	22,000 (1993)	384, 395
Warao	wba	Isolate	Isolate	Venezuela	28,100 (2007)	276, 285–286
Waray	wrz	Australian	Waray	Australia	4 (1981)	344, 377
Wardaman	wrr	Australian	Yangmanic	Australia	50 (1983)	60, 63
Warekena (Guarequena)	gae	Arawakan	Arawakan	Colombia, Brazil, Venezuela	500 (2001)	344, 353, 365
Wari' (Pakaásnovos)	pav	Chapacura-Wanhan	Chapacura-Wanhan	Brazil	1,930 (1998)	60, 67–68, 76, 88, 97, 104–105, 114, 135, 138, 228, 243, 276, 294, 344, 358
Warlpiri	wbp	Australian	Pama-Nyungan	Australia	2,670 (1996)	88, 107
Wayampi	oym	Tupian	Tupi-Guaraní	Brazil	1,180 (2000)	276, 289
Welsh	cym	Indo-European	Celtic	UK	537,870 (1991)	14, 23, 344, 361–362, 384, 402
Wichí (Wichí Lhamtés Güisnay)	mzh	Matacoan	Matacoan	Argentina, Bolivia	15,000 (1999)	88, 97
Wichita	wic	Caddoan	Caddoan	USA	1 (2008)	60, 65, 228, 243
Wik Ngathana	wig	Australian	Pama-Nyungan	Australia	130 (1981)	276, 286
Winnebago (Ho-Chunk)	win	Siouan	Siouan	USA	11 (2004)	60, 83
Wintu	wit	Penutian	Wintuan	USA	5 (1997)	192, 219
Wolaytta	wal	Afro-Asiatic	North Omotic	Ethiopia	1,230,000 (1994)	384, 413
Worora	unp	Australian	Wororan	Australia	20 (1991)	154, 169
Wu, Changzhou (Chinese, Wu (Hangzhou))	wuu	Sino-Tibetan	Chinese	China	77,201,820 (1984)	60, 76
Xokléng	xok	Macro-Ge	Ge-Kaingang	Brazil	760 (1998)	154, 169
Yagua	yad	Peba-Yaguan	Peba-Yaguan	Peru	5690 (2000)	192, 195, 199–201, 228, 252, 255, 264, 269–270
Yana	ynn	Hokan	Yana	USA	extinct	114, 133

Language	ISO	Family	Genus	Location	Speakers	Page
Yanuyuwa	jao	Australian	Pama-Nyungan	Australia	52 (1996)	65
Yemsa	jnj	Afro-Asiatic	North Omotic	Ethiopia	81,600 (1994)	384, 413
Yi (Nuosu)	iii	Sino-Tibetan	Burmese-Lolo	China	2,000,000 (2000)	276, 282, 302, 323
Yiddish (Yiddish, Western)	yih	Indo-European	Germanic	Central Europe	2,255,074	384, 408, 412
Yidiny	yii	Australian	Pama-Nyungan	Australia	12 (1981)	60, 65
Yimas	yee	Lower Sepik-Ramu	Lower Sepik	Papua New Guinea	300 (2000)	60, 72
Yimas-Arafundi Pidgin	---	Pidgin	Yimas-lexified	Papua New Guinea	max 5 speakers, probably extinct (Foley 2013)	14, 26, 114, 141, 145, 154, 163, 276, 286, 306, 344, 367
Yoruba	yor	Niger-Congo	Defoid	Benin, Nigeria	19,380,800 (1993)	60, 78, 88, 97, 306, 327
Yucatec (Maya, Yucatán)	yua	Mayan	Mayan	Mexico	706,000 (1990)	38, 52, 114, 119, 228, 268
Yukaghir, Kolyma (Yukaghir, Southern)	yux	Yukaghir	Yukaghir	Russia	30 (1995)	192, 195–196
Yukulta (Ganggalida)	gcd	Australian	Tangkic	Australian	extinct	228, 253, 268
Yup'ik, Central	esu	Eskimo-Aleut	Eskimo	USA	16,900 (2000)	306, 337
Yurok	yur	Algic	Yurok	USA	12 (2000)	60, 76
Záparo	zro	Zaparoan	Zaparoan	Ecuador	1 (2000)	228, 258
Zapotec, San Bartholomé Zoogocho (Zapotec, Zoogocho (San Bartholomé))	zpq	Oto-Manguean	Zapotecan	Mexico	638 (2000) (Sonnenschein 2004:1)	276, 298
Zapotec, Tilquiapan	zts	Oto-Manguean	Zapotecan	Mexico	7,000 (2007)	60, 80
Zayse (Zaysete)	zay	Afro-Asiatic	North Omotic	Ethiopia	17,800 (1994)	344, 354
Zoque, San Miguel Chimalapa (Zoque, Chimalapa)	zoh	Mixe-Zoque	Mixe-Zoque	Mexico	4,500 (1990)	114, 143
Zulu	zul	Niger-Congo	Bantoid	South Africa	10,349,100 (2006)	60, 74

Glossary

abessive Grammatical case denoting 'without X'.

ability Dynamic event modality denoting capacity on the part of the subject to carry out an action.

ablative Grammatical case denoting 'from X'.

ablaut (also **gradation**; **vowel gradation**) Non-linear morphological marker involving modification of the root vowel.

absolute tense Tense where the event is placed before, after or simultaneous to the speech point on a time line.

absolute universal Universal that holds for every single language, without exceptions.

absolutive clause (also **participial adverbial clause**) Adverbial clause typically used to give general background for the situation expressed in the matrix clause.

absolutive Grammatical case typically used for arguments with the semantic role of patient.

accent Phonological feature relating to the prosodic property of loudness.

Accessibility Hierarchy of Relativization Hierarchy of which syntactic roles in a sentence are more likely to be relativized and how explicit the relativization will be.

accusative (1) Grammatical case typically used for arguments with the syntactic role of object. (2) See **nominative-accusative**.

accusative-focus See **double-oblique**.

actionality See ***Aktionsart***.

additive clause Adverbial clause expressing that one thing happened in addition to another.

additive plural Number value denoting more of the same item.

addressee honorific Linguistic politeness strategy where the form chosen is dependent on the addressee of the utterance.

adessive Grammatical case denoting 'near/by X'.

adjective Part-of-speech denoting qualities or attributes.

adjective clause See **relative clause**.

adjective phrase Phrase with an adjective as its head.

adjoined relative clause Relative clause where the S_{rel} appears outside the NP_{mat} and may be separate from it.

adjunct (also **peripheral participant**) Non-obligatory clause participant.

adposition Part-of-speech typically used to express the relation between the NP it governs and some other element in the clause.

adpositional phrase Phrase with an adposition as its head.

adstratum relationship See **linguistic area**.

Advanced Tongue Root (**ATR**) Type of vowel modification where the tongue is pushed forward.

adverb Part-of-speech which modifies other categories than nouns.

adverb phrase Phrase with an adverb as its head.

adverbial Constituent which functions as an adverb.

adverbial clause Clause which functions as an adjunct to its matrix clause.

adversative coordination Conjunction denoting the antithetical circumstance.

affirmative declarative Speech act used to assert, describe, etc. something.

affix Bound morpheme that does not carry any lexemic information.

affricate Consonant which starts out as a stop and ends as a fricative.

agent Semantic role typically denoting an entity performing an action. Typically implies a high degree of control.

agglutination See **concatenation**.

agreement (also **concord**) Type of linkage of elements where the dependents take a morphological shape corresponding to that of the head.

Aktionsart (also **actionality**; **lexical aspect**; **derivational aspect**) Semantic specification of the inner structure of an event.

alienable possession Type of possession where the possessed item can be transferred. Also sometimes used to denote constructions involving nouns that are only optionally marked as possessed.

allative Grammatical case denoting 'to X'.
allomorph Variant form of morph.
allophone Variant form of phone.
alveolar Coronal produced by raising the tongue to the alveolar ridge.
ambal See **paral**.
ambient clause See **avalent verb**.
analipsis See **forward ellipsis**.
analytic causative Causative construction where a separate verb is used to denote causation.
analytic construction (also **periphrastic construction**) Construction by way of free morphemes.
anaphoric agreement (also **pro-drop**) Construction where verbal agreement is the only reference to the relevant argument in the clause.
animacy Classification of nouns according to how animate or inanimate the entities are.
Animacy Hierarchy (also **Topicality Hierarchy**) Hierarchy postulating that NPs of different animacy levels are treated differently morphosyntactically.
A-not-A-construction See **disjunctive-negative structure**.
antecedent The NP a pronoun refers to.
antepenultimate stress Stress on the third to the last syllable of the word.
anterior Tense category placing an event before a given reference point on a time line.
anticausative (also **middle**; **mediopassive**) Valency reducing operation where the notion of causation is removed, rendering the argument of causer superfluous.
antipassive Valency reducing operation where the patient-like argument becomes optional or suppressed.
apical Coronal where the tip of the tongue is used.
apodosis The matrix of a conditional clause (the "then clause").
apophony (also **stem mutation**) Derivational device involving internal modification of the stem.
applicative Valency increasing operation by way of a verbal marker where an adjunct becomes an argument.
applied object The new object of an applicative clause.
apposition See **parataxis**.
approximant Sonorant formed by causing such a low degree of obstruction to the airflow that no friction is produced.
areal bias Bias to a cross-linguistic sample where area(s) are either over- or underrepresented.
areal feature Feature consistently found in languages of a particular area.
argument (also **core participant**) Obligatory participant of clause.
article Marker, either free or bound, used to indicate whether the NP referred to is identifiable or not.
articulator One of the two main parts of the speech apparatus forming a sound.
aspect Grammatical device specifying the perspective taken on an event.
aspiration A short burst of air accompanying a sound produced.
associates A group of individuals associated with the focal referent in an associative plural construction.
associative plural (also **group plural**) Plural construction denoting a group of unspecified individuals that are in some way associated with a named referent.
assumptive judgement Epistemic propositional modality denoting that the speaker is inferring something on the basis of what is generally known.
asymmetric negation Negative constructions where the structure of the negative declarative differs from that of the affirmative counterpart.
asyndetic coordination (also **juxtaposition**; '**zero strategy**') Coordination of units without any overt linker.
atelic verb A verb where the *Aktionsart* does not imply any inherent end point.
ATR See **Advanced Tongue Root**.
automatic reduplication A construction where an affix obligatorily triggers reduplication but the reduplication itself does not add any meaning to the construction.
auxiliary Semantically more or less empty verbs conveying grammatical information.
avalent verb (also **zero-intransitives**; **ambient clauses**) A verb with a valency of zero.
avoidance language See **bystander honorific**.
back vowel Vowel produced by retracting the tongue.
back-formation Derivational device involving the shortening of a word by taking out an element similar in form to an affix.
backward ellipsis (also **catalipsis**; **right periphery ellipsis**) Ellipsis of the first occurrence of repeated material.
bahuvrihi compound See **exocentric compound**.
balanced language contact Contact situation where the languages involved are on roughly equal footing.

base In reduplication, the original form from which material is copied and with which that copied (reduplicated) material is fused.

basic object The original object of the applicative clause.

behaviour-and-control properties Syntactic processes that are governed by grammatical relations.

benefactive (1) Grammatical case denoting 'for X'. (2) Semantic role typically denoting the entity that benefits from an action.

bibliographical bias Bias to a cross-linguistic sample based on availability of material.

bidirectional implicational universal A prediction that can be reversed.

bilabial Labial produced with both lips.

bisyndetic Coordinating construction with two coordinators.

bivalent verb A verb with a valency of two.

bleaching See **desemanticization**.

blend Derivational device involving two partly truncated words.

bound morpheme Morpheme which does not function as its own word but has to attach to a host.

bound noun See **inherent possession noun**.

branch Group of items more closely connected to each other than to the other items in a tree structure or diagram.

breathy voice vowel Vowel produced with vocal cords held loosely together.

brother-in-law language See **bystander honorific**.

buoy Sign formed by the subordinate hand serving as a conceptual landmark device which guides the discourse as it continues.

bystander honorific (also **avoidance language**; **honorific register**; '**brother-in-law language**'; '**respectful language**'; '**sideways language**') Linguistic politeness where the form chosen is dependent on who is within earshot of the utterance.

calque See **pattern replication**.

cardinal numeral Numeral used to express the number of individuals in a set.

case Grammatical device used to overtly indicate the relationship that the NP has to some other element in the clause.

catalipsis See **backward ellipsis**.

causal clause See **reason clause**.

causal coordination Coordination denoting cause.

causative Valency increasing operation adding an element of causation, thus making the causer an argument.

causee The argument that does something in the caused event in a causative construction.

causer The argument that causes an event in a causative construction.

chain See **cline**.

character See **variable**.

chereme Minimal meaning distinguishing unit in sign languages. The term (now largely abandoned) was meant to mirror 'phoneme'.

cherology The study of how signs are used to mark linguistic contrasts in sign language. The term (now largely abandoned) was meant to mirror 'phonology'.

circumfix Type of bound morpheme involving at least a prefix and a suffix simultaneously.

circumstantial clause Adverbial clause expressing the circumstance by which a particular situation came to be.

classifier (also **noun classifier**) Free and invariant form which assigns a noun to a given category.

clausal negation Negation negating an entire clause giving the negative counterpart to an affirmative declarative.

clause chaining Cosubordinated construction where a chain of clauses are jointly subordinate to a matrix.

clause participant The entity other than the predicate in the clause.

clause reduction See **ellipsis**.

cleft construction A pragmatically marked construction where focus is achieved by expressing the sentence as a relative clause in which a NP and the relativized NP are coreferential.

clefted constituent The NP in a cleft construction that is coreferential with the relativized NP.

click Non-pulmonic consonant formed by the tongue making both a front and a back point of contact in the roof of the mouth, with the middle lowered to form a pocket of air, which produces a click when the tongue is drawn down and back and the air is released.

cline (also **chain**; **pathway**) Series of intermediate and interlacing stages in the process of grammaticalization.

clipping Derivational device involving the shortening of a word.

clitic Bound morpheme which is syntactically independent from its host.

close vowel See **high vowel**.

closed syllable Syllable with a coda.

closed word class Word class where items are not easily added, thus containing a limited set of items.

coarticulated Segments produced at two places of articulation simultaneously.

code-blend When a speaker simultaneously signs and speaks during an utterance.

code-switching When a language user switches language, dialect or language mode within one utterance.

cognitive interjection Interjection indicating the state of knowledge or thoughts of the speaker.

comitative (1) Grammatical case denoting 'with X'. (2) Semantic role typically denoting an entity accompanying someone or something.

comment That part of the sentence which is separate from and contains new information about the topic.

commissive Deontic event modality where the speaker is certifying that the action will take place.

common noun Noun not denoting a specific individual or place.

comparative A construction typically used with adjectives to denote that an entity has a quality to a greater or lesser degree than some other entity.

complement An attribute to an argument.

complement clause A clause which functions as an argument to its matrix clause.

complementizer An overt marker to signal that a unit functions as a complement clause.

completive Aspect category denoting the completion of an event.

complex clause Clause consisting of several clauses.

complex reduplication Reduplication formed by taking material from the base and partly altering it.

compounding The amalgamation of lexemes to form a new lexeme.

conative interjection Interjection directed at an addressee.

concatenation (also **agglutination**) Morphological process of stringing inflectional affixes together linearly on a word.

concatenative marker Bound morphological marker.

concessive clause Adverbial clause which makes a contrast between the matrix and the adverbial clause.

concord See **agreement**.

conditional clause Adverbial clause which expresses a condition.

conjunction Part-of-speech which serves to connect entities. Also used to denote conjunctive coordinations in particular.

conjunctive coordination (also **conjunction**) Conjunction which assigns the connected entities equal status.

consonant Segment formed by creating some obstacle for the air as it passes from the lung through the mouth.

constituent A functional component in a larger structure.

constituent interrogative See **content question**.

constituent negation Negation with scope over only a particular constituent in a clause.

constituent order (also **word order**) The arrangement of functional components.

contact induced grammaticalization Process where a lexical item in the recipient language grammaticalizes to match a grammatical category in the source language.

contact language Cover term for languages which emerged in situations of extreme contact.

content question (also **constituent interrogative**; **information question**; **question-word question**; ***wh*-question**) Question formed with an interrogative phrase and demanding a specific answer containing other information than merely confirmation or nonconfirmation.

content word Word with more or less concrete and specific lexemic meanings.

contour tone Tone where the pitch moves from one level to another over the TBU.

contrastive segment Minimal meaning distinguishing unit.

controller The element which triggers agreement.

convenience sample A sample based on what kind of data is accessible.

converb Non-finite verb with the main function of denoting that a clause is a subordinate adverbial.

convergence area See **linguistic area**.

conversion (also **zero-derivation**) Derivational process where a word changes part-of-speech category without any modification to the word itself.

coordinate compound See **copulative compound**.

coordinating conjunction See **coordinator**.

coordination Process where linguistic units of the same status are linked together.

coordination reduction See **ellipsis**.

coordinator (also **coordinating conjunction**) An overt linker of coordination.

coordinator omission Construction where all but one coordinator are dropped when more than two units are coordinated.

copula Semantically empty formative functioning as a link between an NP and a predicate.

copulative compound (also **coordinate compound**; **dvandva compound**) Compound where the entities in the resulting lexeme have equal status.

core participant See **argument**.

coronal Consonant produced by raising the front part of the tongue towards the roof of the mouth.

correlative clause Relative clause where the NP_{mat} occurs inside the S_{rel} but also has an overt reference in the matrix clause.

cosubordination Complex clause where the units are neither embedded in each other nor independent from each other.

count noun Noun denoting an entity which can be counted.

creaky voice vowel Vowel produced with the vocal cords held tightly together.

creole Language which typically ultimately emerged in a situation of extreme language contact and which has become the mother tongue of its speech community.

creolization (also **nativization**) Process in which a pidgin or jargon is structurally and functionally extended as it becomes the mother tongue of its speech community.

cross-linguistic comparison Comparison across different languages.

cultural bias Bias to a cross-lingusitic sample where particular cultures are over- or underrepresented.

cumulative morpheme (also **polyexponential morpheme**; **portmanteau morpheme**) Morpheme which encodes several pieces of grammatical information simultaneously.

database An organized collection of data.

dative shift Valency transposing operation by which an indirect object is formally realized like a direct object.

dative Grammatical case typically used for arguments with the syntactic role of indirect object.

decategorialization Stage in grammaticalization where an item loses the morphosyntactic properties that identify it with a specific grammatical category.

declarative Type of speech act forming a statement.

deductive judgement Epistemic propositional modality denoting that the speaker is inferring something on the basis of external evidence.

defining relative clause See **restrictive relative clause**.

definite article Marker, either free or bound, used to encode specificity and indicate that the NP referred to is identifiable.

degrammaticalization Reversed grammaticalization whereby a more grammatical item becomes less grammatical.

degree adverb Type of adverb expressing the degree of the quality it modifies.

demonstrative pronoun Pronoun which points something out in a situation or sentence.

dental Coronal produced by raising the tongue to the back of the front teeth.

deontic modality Event modality where external factors initiate or condition potential actions.

dependent An element which is attached or subordinate to a head.

dependent clause See **subordinate clause**.

dependent marking Morphological strategy where the syntactic relations between two elements are marked on the subordinate element.

derivation Morphological process by which new lexemes are created.

derivational affix Affix used to create new lexemes.

derivational aspect See ***Aktionsart***.

desemanticization (also **bleaching**; **emancipation**; **semantic reduction**) Stage in grammaticalization where the sum of the semantic content in a lexical item is reduced to only the grammatical content.

diachronic Temporal dimension concerned with changes over time.

diathesis See **voice**.

diffusion (also **propagation**) The gradual spread of an item.

diffusion area See **linguistic area**.

diphthong Vowel produced with a change in quality during the articulation.

direct causation A causative construction where the link between cause and event is tight.

direct evidential See **sensory evidential**.

direct object Syntactic role typically denoting the 'undergoer', patient or theme of an action.

direct speech act Speech act where the content of the utterance corresponds to that of the speaker's intentions with it.

directive modality Deontic event modality where the speaker tries to initiate an action.

discontinuous reduplication Reduplication where other morphological material may appear between the reduplicant and the base.

discourse A connected series of utterances.

disjunction See **disjunctive coordination**.

disjunctive coordination (also **disjunction**) Coordination which distinguishes between alternatives.

disjunctive-negative structure (also '**A-not-A-construction**') Polar question formed by giving the verb twice, once in its positive and once in its negative form.

displacive language contact Contact situation where one (or several) language(s) dominate(s) over other languages.

ditransitive verb Verb which takes two objects, i.e. has three participants.

dorsal Consonant produced with the body of the tongue.

double headed relative clause Relative clause where the full NP_{mat} is repeated in the S_{rel} in addition to an overt reference in the matrix clause.

double marking (1) Morphological strategy where the syntactic relations between two elements are marked on both the head and the dependent. (2) In language contact: when borrowed material occurs parallel to native material.

double object (also **neutral**) Alignment strategy where the recipient-like and theme-like arguments in a ditransitive clause and the patient-like argument in a monotransitive clause are all encoded the same way.

double-oblique (also '**accusative-focus**') Alignment strategy where the agent-like and patient-like arguments in a transitive clause are coded the same way but the single argument in an intransitive clause is coded differently.

doubly articulated Coarticulated consonant produced simultaneously at two different places but in the same manner.

dual Number value referring to exactly two entities.

dummy subject (also **expletive subject**) Semantically empty element whose role is to fill the syntactic position of a subject argument but that does not refer to any actual entity.

duplifix Type of reduplication where the first consonant of the base is copied and a vowel is inserted between the reduplicant and the base.

duration Prosodic feature yielding e.g. either short or long segments.

durative verb A verb where the *Aktionsart* implies an event with an internal structure made up of a string of phases.

dvandva compound See **copulative compound**.

dynamic modality Event modality where internal factors are what condition or initiate an action.

dynamic verb A verb where the *Aktionsart* implies an inherent element of change.

economy Proposed explanation for language universals referring to (*i*) that frequent elements tend to become reduced in form and (*ii*) that highly predictable elements tend to become eliminated.

ejective Non-pulmonic consonant formed by closing off the glottis and raising the larynx at the time of release, which pushes out a short, sharp burst of air and creates a kind of popping sound.

elicitation Method of collecting data where the investigator asks the consultant specific questions.

ellipsis (also **clause reduction**; **coordination reduction**) Construction where one or more elements have been omitted.

emancipation See **desemanticization**.

embedding Construction where one unit is included as a dependent part of a larger unit.

enclitic (also **postclitic**) Clitic that attaches at the end of its host.

endangered language Language where a dwindling number of speakers learn the language natively.

endocentric compound (also **tatpuruṣa compound**) Compound where the lexeme refers to a subclass of one of the compounded elements.

endoclitic Clitic which places itself inside a root or stem.

enhancement Process in language contact where an already existing tendency in the recipient language is enhanced by the existence of a structure in the source language.

entity sign Sign used for an item or entity, essentially corresponding to the category Noun.

epistemic modality Propositional modality which codes qualitative judgements about the information in a given proposition.

ergative (1) Grammatical case typically used for arguments with the semantic role of agent. (2) See **ergative-absolutive**.

ergative-absolutive (also **ergative**) Alignment strategy where the single argument of an intransitive clause and patient-like argument of a transitive clause are coded the same way but the agent-like argument of the transitive clause is coded differently.

erosion Stage in grammaticalization where a frequently used item loses phonetic substance.

essive Grammatical case denoting 'at/as X'.

event modality Modality pertaining to potential (probable or possible) actions.

event sign Sign used for an action or event, essentially corresponding to the category Verb.

evidential Propositional modality which codes the type of evidence a speaker has for a given proposition.

exclusive Value in pronominal systems specifically excluding reference to the addressee.

exocentric compound (also **bahuvrihi compound**) A compound where the resulting lexeme is not a subclass of any of the compounded elements.

experiencer Semantic role typically referring to an entity getting cognitive stimulus one way or another.

experiential A tense/aspect category denoting that an event took place repeatedly or over a span of time and the experience of those occasions still holds. See also **perfect**.

expletive subject See **dummy subject**.

exponence The overt morphological realization of grammatical information in a particular morpheme.

expressive interjection Interjection indicating the speaker's mental state.

extension (1) Stage in grammaticalization where an item becomes appropriate in a wider set of contexts. (2) In language contact the process where an already existing feature in the recipient language gets extended in its function or use due to contact with the source language.

face The public image a person wishes to maintain.

face saving speech act Speech act where the risk of threatening a person's face is diminished.

face threatening speech act Speech act where the risk of threatening a person's face is increased.

feature See **variable**.

fieldwork The procedure of acquiring linguistic data from language consultants.

finite complement Complement where the verb carries its own TMA inflection and argument agreement.

flap Rhotic formed by striking the tip of the tongue with a sliding motion against the place of articulation.

flexible word order (also **free word order**) System where the constituent order is determined by other criteria than that of syntactic roles.

flexitivity The amount of allomorphy a morphological system has.

flexive markers Paradigmatic set of grammatical markers where the choice of marker is dependent on syntactico-semantic criteria relating to the head.

floating tone Tone not associated with any syllable but which carries morphological information.

focal referent The named referent of an associative plural construction.

focus (also **rheme**) That part of the sentence which constitutes new information about the topic.

force Semantic role typically denoting a non-conscious instigator of action.

formative Morphological marker.

forward ellipsis (also **analipsis**; **gapping**) Ellipsis where the second of the coordinated units is omitted.

fragment buoy Sign language buoy where the non-dominant hand retains a fragment of a two-handed sign as the dominant hand moves on to a succeeding one-handed sign (i.e the non-dominant hand perseverates into the one-handed sign), thus effectively creating a blend.

free morpheme Morpheme which functions alone as its own word.

free relative clause See **headless relative clause**.

free word order See **flexible word order**.

fricative vowel A syllabic fricative.

fricative Obstruent produced with main articulators placed very close together causing obstruction to the airflow without completely closing it off.

front vowel Vowel produced by moving the tongue forward.

fronting Construction where a constituent is moved to sentence initial position.

full reduplication Reduplication where the entire base is copied.

function word (also **grammatical word**) Word with a grammatical function and an abstract, general meaning, or no meaning at all.

functional class (also **functional category**) Word category which consists of function words. Typically a closed word class.

functional reduction Stage in grammaticalization in which grammatical items lose part or all of their grammatical functions.

fusion The degree to which morphological markers attach to a host.

future Tense category placing an event after the moment of speech on a time line.

future perfect A perfect placed after the moment of speech on a time line.

gap Strategy for relative clauses (1) where there is no overt element which is coreferential with the NP_{mat} or (2) where there is neither any overt element which is coreferential with the NP_{mat} nor any overt relativizer.

gapping See **forward ellipsis**; often called gapping since in European languages the deletion of a repeated verb leaves a gap between the pre- and postverbal constituents.

gender Grammatical classification for nouns or substitutes for nouns. See **noun class**.

genealogical affiliation The genetic origin a language has.

genealogical bias See **genetic bias**.

general number Numeral category which does not give any information as to the number of entities involved.

genetic bias (also **genealogical bias**) Bias to a cross-linguistic sample where some language families or genera are either over- or underrepresented.

genitive Grammatical case typically used to mark possession.

genitive phrase See **possessor phrase**.

gesture A voluntary movement of the body (or some part of the body) which carries meaning to the discourse participants.

glide (1) See **semi-vowel**. (2) An audible transition from one sound to another.

global plural See **greater plural**.

gloss The analytical explanation of a unit (word, morpheme, etc.).

glottal Laryngeal consonant produced with the glottis.

goal Semantic role typically denoting an entity expressing the end point in space for a motion.

governee The governed element (dependent) in a construction.

government Linkage between units where the governor determines the morphological shape of the governee, but where the shape of the governee does not carry any semantic information about the governor.

governor The governing element (head) in a construction.

gradation See **ablaut**.

grammatical accommodation In language contact the process by which a recipient language accommodates an existing feature to also cover the functions of a similar sounding feature in the source language.

grammatical agreement (also **lexical argument**) Construction where a free form reference to the relevant argument in the clause is obligatory irrespective of verbal agreement.

grammatical relation (also **syntactic role**) Formal category signalling the syntactic function of an argument in the clause.

grammatical word See **function word**.

grammaticalization Process by which less grammatical forms become more grammatical.

greater plural (also **global plural**) Number value denoting an exceptionally large number of entities.

group plural See **associative plural**.

habitual Aspect category denoting that an event takes place regularly or is true for an extended period of time.

hand configuration Segmental parameter in sign languages made up of handshape and hand orientation.

hand orientation Segmental parameter in sign languages denoting the way the palm faces during the sign.

handshape Segmental parameter in sign languages denoting the shape of the hand during the sign.

head The main unit of a construction.

head marking Morphological strategy where the syntactic relations between two elements are marked on the superordinate element.

headless relative clause (also **free relative clause**) Relative clause where the NP_{mat} is lacking altogether and the reference to the NP_{mat} is in the S_{rel} itself.

hearsay evidential See **quotative**.

heavy syllable Syllable with a long vowel and/or a coda.

hesternal past Past tense specifically denoting that the event took place prior to the day of speaking.

high vowel (also **close vowel**) Vowel produced by a raised tongue.

historic present Morphological present tense used for a past tense event.

hodiernal past Past tense specifically denoting that the event took place on the day of speaking.

hold Sign language segment where none of the individual components which make up the sign change.

honorific Linguistic device for marking social distinction.

honorific register See **bystander honorific**.

horizontal alignment (also **object alignment**) Alignment strategy where the recipient-like and theme-like arguments in a ditransitive clause are coded the

same way but the patient-like argument in a monotransitive clause is coded differently.

iconicity The principle that formal expressions in a language mirror semantic notions.

illative Grammatical case denoting 'into X'.

imperative Type of speech act forming a command. Also used to denote positive imperative in particular.

imperfective Aspectual category denoting that an event is viewed from within its course.

impersonal passive Passive construction lacking an overt lexical subject.

implicational universal (also **restricted universal**; **typological universal**) Preconditioned statement about linguistic phenomena where correlations between features are hypothesized ('if X, then Y').

implosive Non-pulmonic consonant produced by moving the vocal folds downward which causes air to move inward for a short instance.

in situ Lit. 'in place'; used to denote that a constituent remains in the same position in the clause.

inalienable possession Type of possession where the possessed item cannot be transferred. Also sometimes used to denote constructions involving nouns that are obligatorily marked as possessed.

inclusive Value in pronominal systems specifically including reference to the addressee.

incorporation Valency decreasing operation which essentially involves the compounding of the head and one of its arguments.

indefinite article Marker, either free or bound, used to indicate that the NP referred to is not identifiable.

indefinite pronoun Pronoun used to refer to non-specific entities.

indirect causation A causative construction where the link between cause and event is not tight.

indirect evidential See **reported evidential**.

indirect object Syntactic role typically denoting the recipient or 'addressee' of an action.

indirect speech act Speech act where the content of the utterance does not correspond to that of the speaker's intentions with it.

indirective Alignment strategy where the patient-like argument in a monotransitive clause and the theme-like argument in a ditransitive clause are coded the same way but the recipient-like argument in a ditransitive clause is coded differently.

inessive Grammatical case denoting 'inside X'.

inferential evidential Reported evidential where a speaker infers the truth about a proposition based on physical evidence.

infinitive complement Complement clause where the verb is in its infinitive form and carries no agreement marking.

infix Affix which places itself inside a morpheme.

inflectional affix Affix which carries grammatical information.

informant See **language consultant**.

information question See **content question**.

inherent possession noun (also **bound noun**; **obligatorily possessed noun**) Category of nouns where the entity referred to is obligatorily specified for possession.

initial stress Stress located on the first syllable.

innovation New linguistic form.

inposition Adposition which places itself inside the NP it governs.

instrument Semantic role typically denoting the entity used to perform an action.

interjection Part-of-speech which forms a full utterance and is used to express a speaker's mental state or reaction.

interlinearization (also **interlinearized glossing**; **interlinearized morpheme translation**) Morpheme-by-morpheme analysis and glossing from one language to another.

interlinearized glossing See **interlinearization**.

interlinearized morpheme translation See **interlinearization**.

internal relative clause See **internally headed relative clause**.

internally headed relative clause (also **internal relative clause**) Relative clause where the NP_{mat} appears inside the S_{rel} and is not represented in the matrix clause.

International Phonetic Alphabet (**IPA**) Standard notational system for representing the sounds of spoken languages.

interrogative Type of speech act forming a request for information.

interrogative phrase Phrase in content question replacing the constituent asked about.

interrogative pronoun Pronoun used to form content question.

interrogative tag See **tag question**.

Intonation Unit (**IU**) Larger prosodic unit spanning over a phrase, clause or sentence.

intransitive verb Verb that has only one participant (the subject).

introflexive See **root-and-pattern**.

inversion Valency transposing operation where the alignment between the arguments in a clause are exchanged.

IPA See **International Phonetic Alphabet**.

irrealis Mood used to indicate that the speaker cannot assert that a proposition is true.

isolating marker Free morpheme marker.

isomorphism Property where items with similar functions become formally similar.

iterative Aspect denoting the repetition of an event.

IU See **Intonation Unit**.

juxtaposition (1) See **asyndetic coordination**. (2) See **parataxis**.

labial Consonant produced with the lips.

labiodental Labial produced with the upper jaw front teeth and the lower lip.

laminal Coronal where the blade of the tongue is used.

language consultant (also **informant**) A speaker (usually native) of a language acting as source for language data.

language contact Contact between languages.

language description Description and analysis of a language based on available material.

language documentation Collection of raw data of a language.

language family The highest level of a group of affiliated languages.

language genus A level of genealogical language classification which is postulated to be globally comparable in time depth with different language famililes.

language isolate Language without any known relatives.

language shift When a population abandons one language in favour of another.

language universal Property that is hypothesized to hold for all or most known human languages.

lateral Consonant formed by closing off a stretch in the middle of the mouth and letting the air flow on one or both sides of the tongue.

length Prosodic property of duration.

level tone Tone where the pitch stays at the same level over the TBU.

lexeme The minimal semantic unit.

lexical argument See **grammatical agreement**.

lexical aspect See ***Aktionsart***.

lexical category See **lexical class**.

lexical causative Causative where the notion of causation is inherent in the (main) verb itself.

lexical class (also **lexical category**; **content word**) Word category which consists of content words. Typically an open word class.

lexical verb The verb which carries the semantic content of the verbal construction.

lexifier (also **superstrate**) The language from which the majority of the lexicon originates in a contact language.

light syllable Syllable with a short vowel and no coda.

linguistic area (also **adstratum relationship**; **convergence area**; **diffusion area**; ***Sprachbund***) A delimited area containing a group of not closely related languages which share a combination of linguistic features.

linguistic typology (also **typology**) Subdiscipline of linguistics engaging in the systematic study and comparison of language structures.

linked coordination See **syndetic coordination**.

linking adverb (also **text adverb**) Adverb serving to link a sentence to the previous one.

liquid Cover term for rhotic and lateral apico-alveolar approximants.

list buoy Sign language buoy where the non-dominant hand maintains a list of an ordered set.

loan translation See **pattern replication**.

location Segmental parameter in sign languages referring to the location of the signer's hand on the body or around the space of the signer.

locative (1) Grammatical case denoting 'in/at X'. (2) Semantic role typically expressing the point in space for an event or entity.

locative clause Adverbial clause functioning as a place adverb.

long segment Segment produced with long duration.

loudness Prosodic feature yielding e.g. either stressed or unstressed syllables.

low vowel (also **open vowel**) Vowel produced with a lowered tongue.

main clause (also **superordinate clause**) An independent clause. Sometimes called matrix clause.

manner adverb (also **predicate adverb**) Adverb which typically modifies the verb.

manner clause Adverbial clause functioning as manner adverb.

manner of articulation The way an obstacle for the airflow is created in the speech apparatus.

marked nominative Nominative-accusative alignment where the nominative carries an overt marker and the accusative is unmarked.

marker of comparison Marker indicating that one unit is being compared with another.

mass noun (also **noncount noun**) Noun denoting an entity which is not in discrete units that can be counted.

matrix clause A clause which contains an embedded subordinate clause. Sometimes used to denote main clause. See also **reference clause**.

matter borrowing Borrowing where both the phonological shape and the morphological form are taken over from the source language to the recipient language.

medial clause (also **medial verb**) Cosubordinated clause embedded in a matrix clause.

medial monosyndeton Coordination strategy where the coordinator precedes the second coordinated unit.

medial verb See **medial clause**.

mediopassive See **anticausative**.

mensural numeral classifier Entitiy which provides a mass nouns with a countable unit.

mesoclitic Clitic which attaches itself between the host and the inflectional affixes of that host.

middle See **anticausative**.

minimal pair Set of two words where only one phoneme in the same position differs.

minimal set Set of more than two words where only one phoneme in the same position differs.

mixed language Language which typically emerged in a situation of stable bilingualism and thus typically has two ancestors.

modality (1) Grammatical category coding the attitude of the speaker towards a given proposition. (2) When differentiated from mood: the semantic label of attitudes towards events. See also **mood**.

mode Cover term for mood and modality.

monoexponential morpheme See **separative morpheme**.

monophthong Vowel produced without any change in quality during the articulation.

monosyndetic Coordinating construction with only one coordinator.

monotransitive verb See **transitive verb**.

monovalent verb A verb with a valency of one.

mood (1) Grammatical category coding the attitude of the speaker towards a given proposition. (2) When differentiated from modality: the higher-level distinction for the whole clause of realis/irrealis. See also **modality**.

moribund language A language no longer learned natively, thus destined to go extinct when the last generation of speakers dies.

morph The smallest discrete unit which conveys some kind of meaning.

morpheme The abstract notion of the smallest meaning carrying unit, realized by any one of a given set of morphs.

morphological causative Causative construction where a morphological process is applied to the base verb in order to get the notion of causation.

morphology The study of shapes, specifically how minimal units of information combine to form words.

movement Sign language segment where one or more of the individual components which make up the sign change.

nasal Sonorant produced by letting air flow through the nasal passage during the articulation.

nasalized vowel Vowel produced with the soft palate lowered, thereby letting air through the nasal as well as oral passage.

nativization See **creolization**.

negation Device used to reverse the truth of a proposition.

negative adverb Adverb negating a clause.

negative declarative Speech act used to assert that something is not true.

negative face That public self-image of a person which maintains his/her independence and freedom from imposition.

negative imperative (also **prohibitive**) Imperative used to command the addressee not to perform an action.

negative pronoun Pronoun referring to a non-existing entity.

neutral (1) Alignment strategy where the agent-like and patient-like arguments in a transitive clause and the single argument in an intransitive clause are all encoded the same way; (2) See **double object**.

node Connecting point of two or more branches in a tree structure or diagram.

nominalized complement Complement clause where the verb is made to function as a noun.

nominative Grammatical case typically used for arguments with the syntactic role of subject.

nominative-accusative (also **accusative**) Alignment strategy where the single argument of an intransitive clause and agent-like argument of a transitive clause are coded the same way but the patient-like argument of the transitive clause is coded differently.

noncount noun See **mass noun**.

non-defining relative clause See **non-restrictive relative clause**.

non-dynamic verb See **stative verb**.

non-finite complement Complement clause where the verb is not in a finite form.

nonflexive marker Morphologically invariant grammatical marker.

nonfuture Tense specifically denoting that an event does not take place after the speech moment.

non-linear marker Morphological marker which involves some kind of modification to the host.

non-manual Suprasegmental marker in sign languages made by other means than with the hand, e.g. facial expressions and body posture.

nonpast Tense specifically denoting that an event did not take place before the speech moment.

non-pulmonic Speech sound which does not involve pushing air out of the lungs.

non-reduction Relative clause strategy where the NP_{mat} reappears as a full NP in the S_{rel}.

non-restrictive relative clause (also **non-defining relative clause**) Relative clause which only adds information about the NP_{mat} but does not restrict the potential referents to any specific NP out of several.

noun Part-of-speech typically referring to concrete or abstract entities.

noun class Grammatical classification for nouns or substitutes for nouns. See **gender**.

noun classifier See **classifier**.

noun incorporation Valency reducing operation whereby an argument is incorporated into the verb.

noun phrase (**NP**) The entity which functions as an argument and has a noun or pronoun as its head.

NP See **noun phrase**.

NP_{mat} (For NPmatrix) The head NP in a relative clause.

NP_{rel} See **relativized NP**.

number Grammatical device for expressing how many real world entities are being referred to.

numeral Part-of-speech typically used to indicate the number of something.

numeral classifier See **sortal numeral classifier**.

object Syntactic role typically denoting that core argument of a clause which is not a subject, traditionally defined as the 'recipient' or 'undergoer' of an action.

object alignment See **horizontal alignment**.

object complement Complement clause which functions as the object of the matrix clause.

obligative Deontic event modality expressing obligations.

obligatorily possessed noun See **inherent possession noun**.

Observer's Paradox The paradoxical problem that the mere presence of a linguist will affect the language s/he is investigating.

obstruent Consonant produced with a high degree of friction.

obviative Morphological third person form marking for the lesser focussed argument in a clause.

open syllable Syllable lacking a coda.

open vowel See **low vowel**.

open word class Word class where items are easily added, thus containing a potentially unlimited set of items.

oral vowel Vowel produced with the soft palate raised, thereby closing the nasal passage and letting out the air through the oral passage. Also referred to as **vowel**.

ordinal numeral Numeral used to express the rank in a series.

overt coding properties Overt features of the grammatical code.

palatal Dorsal formed by raising the body of the tongue to the palate.

paradigmatic feature Feature belonging to a mutually exclusive set.

parafix Type of bound morpheme involving at least two types of affixes simultaneously.

paral (also **ambal**) Number value used with naturally occurring pairs, such as eyes or ears.

parameter See **variable**.

paratactic relative clause A relative clause optionally containing a full NP_{mat}; the relative clause is not overtly subordinate and is only loosely joined with the matrix clause.

parataxis (also **apposition**; **juxtaposition**) Juxtaposition of a complement and matrix clause without any overt complementizer.

partial reduplication Reduplication where only a set part of the base is copied.

participant role The syntactico-semantic role that an argument has in the clause.

participial adverbial clause See **absolutive clause**.

participial complement Complement clause with the verb in its participial form. Typically functions as a modifier rather than a constituent.

part-of-speech (also **word class**) The grammatical category of a word.

passive Valency reducing operation where the agent-like argument of a transitive clause gets demoted or omitted while the patient-like argument gets promoted to the syntactic role of subject.

past Tense category specifying that an event took place before the moment of speech.

past perfect (also **pluperfect**) A perfect placed before the moment of speech.

pathway See **cline**.

patient Semantic role typically denoting the entity affected by an action. Typically implies little or no control.

pattern replication (also **calque**; **loan translation**) Borrowing where the function but not the form of a linguistic feature is taken over from the source language to the recipient language.

paucal Number value referring to a small group of entities.

penultimate stress Stress on the second to the last syllable.

perfect A tense/aspect category used to describe an event which took place before a given reference point but which is still relevant at that point. Sometimes analysed as having both resultative and experiential use.

perfective Aspectual category denoting that an event is viewed as a bounded whole.

peripheral participant See **adjunct**.

periphrastic construction See **analytic construction**.

permissive Deontic event modality expressing permission.

personal passive Passive construction containing an overt lexical subject.

personal pronoun Pronoun referring to the speaker(s), the addressee(s) and/or contextually unambiguous entities.

pharyngeal Radical consonant produced by moving the root of the tongue towards the pharynx.

pharyngealized vowel Vowel produced with the tongue root retracting.

phatic communion Communication with the primary function of maintaining social contact (rather than to convey information).

phatic interjection Interjection used as a communicative cue.

phone The smallest discrete unit which forms a contrastive segment.

phoneme The abstract notion of the smallest meaning distinguishing unit, realized by any one of a given set of phones.

phonetics The study of the physical characteristics of speech.

phonology The study of the abstract linguistic patterning of minimally contrastive segments.

phrase A word or group of words which form a constituent and function as a single unit within a larger syntactic structure.

pidgin Language which typically emerged in a situation of extreme language contact and which is not the mother tongue of its speakers.

pied-piping Syntactic strategy where the constituents of an adpositional phrase remain in their conventional order, e.g. in subordinate clauses.

pitch Prosodic feature yielding different tones.

place of articulation The place where an obstacle for the airflow is created in the speech apparatus.

plosive (also **stop**) Obstruent consonant where the airflow has been completely closed off.

pluperfect See **past perfect**.

plural Number value referring to more than one entity.

pointer buoy Sign language buoy where the non-dominant hand points towards an important element in the discourse.

polar question (also **yes-no question**) Interrogatives where the answer is typically expected to either confirm or disconfirm the proposition.

polarity Opposition between two options, e.g. affirmative and negative.

politeness A mode of interaction which shows awareness and respect for someone else's face.

polyexponential morpheme See **cumulative morpheme**.

polysynthesis Morphological process which typically allows the fusion of more than one lexeme.

portmanteau morpheme See **cumulative morpheme**.

position Segmental parameter in sign languages denoting the position of the hands during the sign.

positive The form of an adjective not marked for any degree of comparison.

positive face That public self-image of a person which maintains his/her acceptance and group belonging.

positive imperative (also **imperative**) Imperative used to command an addressee to perform an action.

possessee (also **possessum**) The head noun of a genitive construction.

possessor phrase (also **genitive phrase**) The phrase in a genitive construction which describes who or what the head noun belongs to.

possessum See **possessee**.

post-alveolar Coronal produced by raising the tongue to the back of the alveolar ridge.

postaspirated Segment produced with a short burst of air immediately after it has been articulated.

postclitic See **enclitic**.

posterior Tense category placing an event after a given reference point on a time line.

postnominal relative clause Relative clause where the S_{rel} follows the NP_{mat}.

postposing With subject complements: strategy of moving the subject complement clause to after the matrix clause; in some languages the empty subject slot is filled with a pronoun.

postposition Adposition which follows the NP it governs.

postpositional phrase Phrase with a postposition as its head.

postpositive coordinator Coordinating construction where the coordinator follows the units it is coordinating.

pragmatic inference Communicative situation where speaker and hearer presuppose capacity to draw conclusions beyond the semantic meaning of the utterance.

pragmatic role Formal category describing the status of the information given in a clause.

preaspirated Segment produced with a short burst of air immediately before it is articulated.

predicate That part of a clause or sentence which asserts something about the subject.

predicate adverb See **manner adverb**.

prefix Affix which places itself at the beginning of its host.

prenominal relative clause Relative clause where the S_{rel} precedes the NP_{mat}.

preposition Adposition which precedes the NP it governs.

preposition stranding Syntactic strategy where in a relative clause the NP_{rel} is fronted and the preposition is left hanging at the end of the clause.

prepositional phrase Phrase with a preposition as its head.

prepositive coordinator Coordinating construction where the coordinator precedes the units it is coordinating.

present Tense category specifying that an event is taking place at the moment of speech.

primary compound See **root compound**.

primary topic The central topic in a sentence with more than one topic.

probabilistic universal See **statistical universal**.

probability sample Sample needed to check for statistical tendencies and correlations of features, where the variables are set beforehand and mapped according to their presence or absence in a given language.

processing The cognitive effort it takes to comprehend linguistic structures.

proclitic Clitic that attaches itself at the beginning of its host.

pro-drop See **anaphoric agreement**.

progressive Aspect category denoting that an event is ongoing.

prohibitive See **negative imperative**.

pronominal gender system System where the gender of the free pronoun is determined by its antecedent.

pronoun Part-of-speech typically used to substitute a noun or NP.

pronoun avoidance Politeness strategy where the speaker avoids using the pronoun to refer to someone.

pronoun retention Relative clause strategy where the S_{rel} contains a resumptive pronoun which is coreferential with the NP_{mat}.

propagation See **diffusion**.

proper name (also **proper noun**) Noun referring to a specific individual or place.

proper noun See **proper name**.

property sign Sign used as a modifier of either entity or event signs.

propositional modality Modality which codes the speaker's attitude towards the truth value of the information given in the proposition.

prosodic feature See **suprasegmental feature**.

prosodic formative See **suprasegmental feature**.

prosodic modification Derivational device involving stress or tone.

prosody Variations in duration, pitch, loudness and rhythm.

protasis The subordinate of a conditional clause (the "if clause").

prototype The most representative example of something.

proximate Morphological third person form marking for the more focussed argument in a clause.

pulmonic Speech sound which involves pushing air out of the lungs.

punctual verb A verb where the *Aktionsart* implies that there is no internal structure to the event.

purpose (also **purposive**) Semantic role typically denoting the entity which is the reason for an action.

purpose clause Adverbial clause which expresses the purpose for an action.

purposive See **purpose**.

question particle Free or bound morpheme used to turn a declarative into an interrogative.

questionnaire Written form of data elicitation.

question-word question See **content question**.

quotative (also **hearsay evidential**; **reportative**; **second-hand evidential**) Reported evidential where a speaker indicates that s/he has been told about an event.

radical Consonant produced with the root of the tongue.

realis Mood used to indicate that the speaker is very sure that a proposition is true.

reality conditional Adverbial clause denoting a real condition.

reason clause (also **causal clause**) Adverbial clause which explains why an event took place.

recipient Semantic role typically denoting an entity receiving something.

recipient language A language which adopts a linguistic feature from some other language.

reciprocal pronoun Pronoun which is co-referential with a co-occurring nominal and which expresses mutuality.

reduplicant The repeated element in reduplicated constructions.

reduplication Morphological process where a set amount of phonological material is copied from a base form and fused with it to form a stem.

reference clause Sometimes used to specify the matrix clause of medial clauses.

reference point Point on a time line before, after or simultaneous to which an event occurs.

referent honorific Linguistic politeness where the form chosen is dependent on what or whom is being referred to.

reflexive pronoun Pronoun which is co-referential with a co-occurring nominal.

reinterpretation In language contact the process where borrowed matter acquires a new function in the recipient language.

relative clause (also **adjective clause**) Clause which modifies an NP by delineating a specific antecedent to which a certain proposition is true.

relative pronoun Pronoun used to introduce an S_{rel}.

relative tense Tense where an event is placed before, after or simultaneous to a given reference point on a time line.

relativized NP (also $\mathbf{NP_{rel}}$) An overt element which is coreferential with the NP_{mat}.

relativizer Morphological marker indicating a relative clause.

remoteness Degrees of distance from the speech point on a time line.

replacement (also **substitution**) Morphological process where a regular marker replaces a part of the stem.

reportative See **quotative**.

reported evidential (also **indirect evidential**) Evidential where a speaker indicates that the evidence had for a proposition is through hearsay or similar sources.

respectful language See **bystander honorific**.

restricted universal See **implicational universal**.

restricting clause See $\mathbf{S_{rel}}$.

restrictive relative clause (also **defining relative clause**) Relative clause which restricts the potential referents in the matrix clause to a specific NP out of several.

resultative A tense/aspect category used to describe an event which took place before a given reference point but which is still relevant at that point. See also **perfect**.

retroflex (also **sub-apical**) Coronal where the underside of the tip of the tongue is used.

rheme See **focus**.

rhotacized vowel See **rhotic vowel**.

rhotic Cover term for any liquid approximant which forms an "r-sound".

rhotic vowel (also **rhotacized vowel**) Vowel produced with "r-colour" modification.

rhyme Syllable nucleus plus syllable coda.

rhythm The pattern of how stressed and unstressed syllables alternate.

right periphery ellipsis See **backward ellipsis**; often called right periphery ellipsis since in European languages it is typically the last element of the first coordinated clause that is deleted.

root the smallest lexemic unit.

root compound (also **primary compound**) A compound formed with root morphemes, without any affixes and where none of the entities functions as an argument to the other.

root-and-pattern (also **introflexive**) Morphological strategy where grammatical information is conveyed through the insertion of a pattern of vowels into a consonantal root.

round vowel Vowel produced with rounded lips.

scope That section which is affected by the meaning of a particular form.

second stress Stress located on the second syllable.

secondary topic The topic which stands in some relation to the primary topic in a sentence with more than one topic.

second-hand evidential See **quotative**.

secundative Alignment strategy where the patient-like argument in a monotransitive clause and the recipient-like argument in a ditransitive clause are coded the same way but the theme-like argument in the ditransitive clause is coded differently.

segment An element which forms a sequence with other segments.

semantic reduction See **desemanticization**.

semantic role (also **thematic role**; **theta role**) Formal category denoting the roles that participants play in a given situation.

semantic shift A change in the meaning of words.

semantic valency The amount of necessary participants that are inherently specified in the meaning of a verb.

semi-vowel (also **glide**) A segment which is like a vowel but which may take the place of a consonant in the syllable structure.

sensory evidential (also **direct evidential**) Evidential where a speaker indicates that the evidence had for a proposition is through evidences of the senses (e.g. auditory or visual).

sentence Largest independent unit over which grammatical rules operate.

sentence adverb Adverb which modifies an entire sentence.

sentence-like complement (**s-like**) Finite complement which looks similar in form to what it would have looked like if it had been a main clause.

separative morpheme (also **monoexponential morpheme**) Morpheme which only encodes one single piece of grammatical information at the same time.

serial verb See **serial verb construction**.

serial verb construction (also **serial verb**) Cosubordinate construction where a string of verbs act together as a single predicate.

serial verbs of motion Serial verb construction in sign languages specifically involving motion signs.

setting adverb Adverb specifying either space or time.

short segment Segment produced with short duration.

sideways language See **bystander honorific**.

sign language Language produced by way of using the hands, upper body and face to produce signs.

simple clause Independent clause not containing other clauses as embedded elements.

simple reduplication Reduplication where the reduplicant repeats material from the base without altering it.

simultaneous clause Adverbial clause which indicates that two events overlap.

singular Number value referring to exactly one entity.

situational modality Cover term for deontic obligations and permission as well as dynamic abilities.

s-like See **sentence like complement**.

sonorant Consonant where the air flows relatively freely through the speech apparatus.

sonority principle Principle by which the less sonorous a segment is, the further away it is from the core of the syllable.

sortal numeral classifier (also **numeral classifier**) Classifier assigning a noun to a semantic category, irrespective of whether the noun is countable or not.

source Semantic role typically denoting an entity expressing the point of origin for a motion.

source language A language from which a linguistic feature originates.

speculative judgement Epistemic propositional modality denoting that the speaker is uncertain about the factual status of the proposition.

speech act An act performed by making an utterance, as in using an utterance for any of the core purposes of asserting something, asking something or commanding something.

speech point Point on a time line, typically corresponding to the present moment, before, after or simultaneous to which an event occurs.

split ergativity Split system where the alignment system is determined by semantic and/or pragmatic factors in the transitive clause, e.g. the animacy of the participants or the TMA of the verb.

split intransitivity Split system where, depending on its semantic role, the single argument of an intransitive clause is either aligned with the most agent-like or most patient-like argument of a transitive clause.

split system Alignment system with different groupings depending on different factors.

spoken language Language produced by way of using the lungs, vocal cords and oral/nasal cavities to produce sounds.

Sprachbund See **linguistic area**.

S_{rel} (also **restricting clause**) That part of the clause which functions as a relative clause.

stable language A language vigorously used by its community with no discernible sign of losing native speakers.

standard of comparison The unit with which another unit is being compared.

statistical universal (also **probabilistic universal**) Universal which holds for most languages.

stative verb (also **non-dynamic verb**) A verb where the *Aktionsart* implies that there is no element of change, merely a constant state.

stem The base for an inflected word form.

stem mutation See **apophony**.

stop See **plosive**.

stress Phonetic parameter of loudness.

strident vowel Vowel produced with a constriction at the root of the glottis.

sub-apical See **retroflex**.

subject complement Complement clause which functions as the subject of the matrix clause.

subject Syntactic role typically used for the central argument of a clause, traditionally defined as the 'doer' of an action.

subordinated clause (also **dependent clause**) A clause which does not function independently but is embedded in a main clause.

subordinating conjunction Conjunction which makes one entity subordinate to another.

substitution See **replacement**.

substitutive clause Adverbial clause which expresses that an expected event is replaced by another event.

substrate The language(s) which typically are or were the mother tongue(s) of the dominated speaker community using a contact language.

subtraction Morphological process where grammatical information is conveyed by taking out an element of the stem.

suffix Affix which places itself at the end of its host.

superlative Construction typically used with adjectives to denote that an entity is singled out as having a quality to the highest degree.

superordinate clause See **main clause**.

superstrate See **lexifier**.

suppletion Morphological process where a root or stem is paradigmatically replaced with a root or stem of a different etymological origin.

suprasegmental feature (also **prosodic feature**; **prosodic formative**) Contrastive feature which may carry over across segments.

switch reference Referent tracking device indicating whether the subject in the following clause is the same as or different from the present clause.

syllable Unit of speech sound(s) which can be produced in isolation.

syllable coda The part of the syllable following the nucleus.

syllable nucleus The obligatory core of a syllable, typically consisting of either a vowel or some other sonorous segment.

syllable onset The part of the syllable preceding the nucleus.

syllable weight Level of syllabic prominence.

symmetric negation Negative constructions where the structure of the negative declarative is identical to that of the affirmative counterpart except for the presence of the negator.

synchronic Temporal dimension concerned with a single moment in time.

syndetic coordination (also **linked coordination**) Coordination of units by means of an overt linker.

syntactic compound (also **verbal compound**) Compound where the head is a verb and the modifier could have functioned as an argument to that verb.

syntactic pivot The argument around which the rest of the sentence revolves.

syntactic role See **grammatical relation**.

syntactic valency The amount of participants that are present in the clause.

syntax The study of how constituents are arranged.

synthesis A scale indicating how much accumulated information a word can hold.

synthetic construction Construction by way of affixation.

T/V pronoun Pronominal binary distinction used as referent honorific.

taboo relative A relative in whose presence bystander honorifics must be used.

tag question (also **interrogative tag**) A question particle added to a statement in order to get confirmation or disconfirmation of that statement, but which typically also carries with it an implied expectation as to the answer.

tap Rhotic formed by flicking the tip of the tongue with the shortest possible contact against the place of articulation.

target The element whose morphological form is determined by the controller.

target language The language in a displacive contact situation which people aim to learn.

tatpuruṣa compound See **endocentric compound**.

TBU See **Tone Bearing Unit**.

telic verb A verb where the *Aktionsart* implies an inherent end point.

temporal Semantic role typically expressing the point in time for an action or event.

tense Grammatical device which places an event on a time line.

text As data source: (1) any coherent stretch of recorded (and transcribed) language; (2) a written text.

text adverb See **linking adverb**.

thematic role See **semantic role**.

theme (1) See **topic**. (2) Semantic role typically denoting an entity changing location or to which a location is assigned.

theta role See **semantic role**.

third stress Stress located on the third syllable.

time clause Adverbial clause functioning as a time adverbial.

tone Prosodic property of pitch variation.

Tone Bearing Unit (**TBU**) The unit with which a tone is associated, usually the syllable.

topic (also **theme**) The central part of the sentence which is already known in the discourse and which the rest of the sentence is about.

Topicality Hierarchy See **Animacy Hierarchy**.

transference The process where a linguistic form is transferred wholesale from one language into another.

transitive verb (also **monotransitive verb**) Verb which takes only one object, i.e has two participants.

transitivity Grammatical category specifying whether or not a verb takes an object.

translation The closest idiomatic equivalence in language B of an expression in language A. Not to be confused with linguistic analysis.

trial Number value referring to exactly three entities.

trill Rhotic where the tongue is rolled against the place of articulation.

tripartite (1) Alignment strategy where the single argument in an intransitive clause and the agent-like and patient-like arguments in a transitive clause are all coded differently. (2) Alignment strategy where the patient-like argument in a monotransitive clause and the recipient-like and theme-like arguments in a ditransitive clause are all coded differently.

triphthong Vowel produced with two changes in quality during the articulation.

truncation Derivational device involving the shortening of a word, alternatively replacing a suffix with a shorter one.

typological bias Bias to a cross-linguistic sample where linguistic type(s) are either over- or underrepresented.

typological universal See **implicational universal**.

typology See **linguistic typology**.

ultimate stress Stress on the last syllable of the word.

unidirectional implicational universal A prediction that only holds one way, i.e. cannot be reversed.

unidirectionality Hypothesis that grammaticalization moves from less to more grammatical but not the other way about.

unreality conditional Adverbial clause denoting an unreal condition, such as a predicted or imagined situation.

unrestricted universal Statement about independent linguistic phenomena without preconditions.

unrounded vowel Vowel produced without rounding of the lips or with the lips spread.

uvular Dorsal produced by raising the back of the tongue to the uvula.

valency Device specifying how many arguments a verb obligatorily takes.

variable (also **character**; **feature**; **parameter**) Characteristic which may take any one of a specified set of values.

variety sample Sample needed to explore the diversity and distribution of a feature, the aim being to capture as many as possible of the feature values.

velar Dorsal formed by raising the body of the tongue to the velum.

verb Part-of-speech typically referring to actions and processes.

verb agreement (also **verb coding**; **verb concord**) Morphological device for marking participant reference on the verb.

verb coding See **verb agreement**.

verb concord See **verb agreement**.

verbal compound See **syntactic compound**.

voice (also **diathesis**) Grammatical category which announces the semantic role of the subject in the clause.

voiced Segment produced with the vocal folds vibrating.

voiceless Segment produced with no vibration to the vocal folds.

volitive Dynamic event modality denoting willingness on the part of the subject to carry out an action.

vowel Segment formed by letting air flow freely from the lungs through the mouth.

vowel gradation See **ablaut**.

vowel quality The character of the vowel as it is determined by the posture of the articulators.

vowel quantity The duration of the vowel.

***wh*-question** See **content question**.

word class See **part-of-speech**.

word order See **constituent order**.

word-formation The process of creating new words.

yes-no question See **polar question**.

zero strategy See **asyndetic coordination**.

zero-derivation See **conversion**.

zero-intransitive See **avalent verb**.

References

Acharya, J. 1991. *A Descriptive Grammar of Nepali and an Analyzed Corpus.* Washington DC: Georgetown University Press.

Adelaar, W. & Muysken, P. 2004. *The Languages of the Andes.* Cambridge: CUP.

Ahlgren, I. & Bergman, B. 2006. Det svenska teckenspråket. In *Teckenspråk och teckenspråkiga Kunskaps- och forskningsöversikt. SOU 2006* 29: 11–71.

Aikhenvald, A. 2000. *Classifiers. A Typology of Noun Categorization Devices.* Oxford: OUP.

Aikhenvald, A. 2003. Mechanisms of change in areal diffusion. New morphology and language contact. *Journal of Linguistics* 39: 1–29.

Aikhenvald, A. 2004. *Evidentiality.* Oxford: OUP.

Aikhenvald, A. 2006a. Grammars in contact. A cross-linguistic perspective. In Aikhenvald & Dixon (eds), (a) 1–66.

Aikhenvald, A. 2006b. Semantics and pragmatics of grammatical relations in the Vaupés linguistic area. In Aikhenvald & Dixon (eds), (a) 237–66.

Aikhenvald, A. 2006c. Serial verb constructions in typological perspective. In Aikhenvald & Dixon (eds), (b) 1–68.

Aikhenvald, A. 2007. Typological distinctions in word-formation. In Shopen (ed.), Vol. III: 1–65.

Aikhenvald, A. 2010. *Imperatives and Commands.* Oxford: OUP.

Aikhenvald, A. & Dixon, R. M. W. 1998. Evidentials and areal typology. A case study from Amazonia. *Language Sciences* 20(3): 241–57.

Aikhenvald, A. & Dixon, R. M. W. (eds). 2001. *Areal Diffusion and Genetic Inheritance. Problems in Comparative Linguistics.* Oxford: OUP.

Aikhenvald, A. & Dixon, R. M. W. (eds). 2006a. *Grammars in Contact. A Cross-linguistic Typology.* Oxford: OUP.

Aikhenvald, A. & Dixon, R. M. W. (eds). 2006b. *Serial Verb Constructions. A Cross-Linguistic Typology.* Oxford: OUP.

Ameka, F. 2006. Interjections. In Brown (ed.), 743–746.

Ameka, F. 2007. Grammatical borrowing in Likpe (Sɛkpɛlé). In Matras & Sakel (eds), 107–22.

Andersen, T. 1988. Ergativity in Päri. A Nilotic OVS language. *Lingua* 75: 289–324.

Andersen, T. 2002. Case inflection and nominal head marking in Dinka. *Journal of African Languages and Linguistics* 23: 1–30.

Anderson, D. 1962. *Conversational Ticuna.* Yarinacocha: Instituto Lingüístico de Verano.

Anderson, G. 2009. *Auxiliary Verb Constructions.* Oxford: OUP.

Anderson, J. 2000. Auxiliary. In Booij et al. (eds), 820–31.

Anderson, S. 2005. *Aspects of the Theory of Clitics.* Oxford: OUP.

Anderson, V. & Otsuka, Y. 2006. The phonetics and phonology of 'definitive accent' in Tongan. *Oceanic Linguistics* 45(1): 21–42.

Andrews, A. 2007a. Relative clauses. In Shopen (ed.), Vol. II: 206–36.

Andrews, A. 2007b. The major functions of the noun phrase. In Shopen (ed.), Vol. I: 132–223.

Anonby, S. 2009. Language use on the Tubarão-Latundê Reserve, Rondônia, Brazil. *SIL Electronic Language Reports*. 2009-005. 12 pages. <http://www.sil.org/silesr/2009/silesr2009-005.pdf>

Ansaldo, U. & Lim, L. 2013. Singlish language structure dataset. In *Atlas of Pidgin and Creole Language Structures Online,* S. Michaelis, P. Maurer, M. Haspelmath & M. Huber (eds) and the APiCS Consortium (contr.). München: MPDL.

Anward, J. 2006. Word classes/Parts of speech. Overview. In Brown (ed.), 628–632.

Applegate, J. 1955. Shilha. A Descriptive Grammar With Vocabulary and Texts. PhD dissertation, University of Pennsylvania.

Arensen, J. 1982. *Murle Grammar.* Juba: SIL.

Aronoff, M., Meir, I., Padden, C. & Sandler, W. 2003. Classifier constructions and morphology in two Sign Languages. In *Perspectives on Classifier Constructions in Sign Language,* K. Emmorey (ed.), 53–84. Mahwah NJ: Lawrence Erlbaum Associates.

Aronoff, M., Meir, I., Padden, C. & Sandler, W. 2004. Morphological universals and the sign language type. In *Yearbook of Morphology,* G. E. Booij & J. van Marle (eds), 19–39. Dordrecht: Kluwer.

Aronoff, M., Meir, I., Padden, C. & Sandler, W. 2008. The roots of linguistic organization in a new language. *Interaction Studies* 9(1): 133–53.

Asher, R. & Kumari, T. 1997. *Malayalam.* London: Routledge.

Aye, K. 2013. Singapore Bazaar Malay. In Michaelis et al. (eds), Vol. III.

Ayuninjam, F. 1998. *A Reference Grammar of Mbili.* Lanham MD: University Press of America.

Azevedo, M. 2005. *Portuguese. A Linguistic Introduction.* Cambridge: CUP.

Baker, M. 1988. *Incorporation. A Theory of Grammatical Function Changing.* Chicago IL: University of Chicago Press.

Baker, M. 2001. *The Atoms of Language.* New York NY: Basic Books.

Baker, M. 2004. *Lexical Categories. Verbs, Nouns, and Adjectives.* Cambridge: CUP.

Baker, M. 2005. On verb-initial and verb-final word orders in Lokaa. *Journal of African Languages and Linguistics* 26: 125–64.

Baker-Shenk, C. & Cokely, D. 1991. *American Sign Language. A Teacher's Resource Text on Grammar and Culture.* Washington DC: Gallaudet University Press.

Bakker, D. 2010. Language sampling. In Song (ed.), 100–127.

Bakker, P. 2013. Michif. In Michaelis et al. (eds). (Vol. III)

Bamgbose, A. 1966. *A Grammar of Yoruba.* Cambridge: CUP.

Baptista, M. 2013. Cape Verdean Creole of Brava. In Michaelis et al. (eds), Vol. II.

Bauer, L. 2003. *Introducing Linguistic Morphology,* 2nd edn. Edinburgh: EUP.

Bauer, L. 2004. *A Glossary of Morphology.* Edinburgh: EUP.

Beard, R. 1998. Derivation. In Spencer & Zwicky (eds), 44–65.

Beauvoir, J. 1915. *Los selknam. Indígenas de la Tierra del Fuego. Sus tradiciones, consstumbres y lengua.* Buenos Aires: Misiones Salesianas.

Beavon, K. 1986. Anaphora, pronouns and reference in Konzime. In Wiesemann (ed.), 167–90.

Bennis, H. & Haegeman, L. 1984. On the status of agreement and relative clauses in West-Flemish. In *Sentential Complementation. Proceedings of the International Conference Held at UFSAL, Brussels, June 1983,* W. de Geest & Y. Putseys (eds), 33–53. Dordrecht: Foris.

Berchem, J. 1991. *Referenzgrammatik des Somali.* Cologne: OMIMEE Intercultural Publ.

Berghäll, L. 2010. Mauwake Reference Grammar. PhD dissertation, University of Helsinki.

Bergman, B. & Dahl, Ö. 1994. Ideophones in Sign Language? The place of reduplication in the tense-aspect system of Swedish Sign Language. In *Tense, Aspect and Action. Empirical and Theoretical Contributions to Language Typology,* C. Bache (ed.), 397–422. Berlin: Mouton de Gruyter.

Beyer, S. 1992. *The Classical Tibetan Language.* Alany NY: State University of New York Press.
Bhaskararao, P. & Ladefoged, P. 1991. Two types of voiceless nasals. *Journal of the International Phonetic Association* 21(2): 80–88.
Bhat, D. & Pustet, R. 2000. Adjective. In Booij et al. (eds), 757–69.
Bhat, D. 2004. *Pronouns.* Oxford: OUP.
Bickel, B. 2006. What favors rara? A Himalayan case study. Paper presented at the Rara and Rarissima conference, Leipzig, March 29–April 1.
Bickel, B. 2007. Typology in the 21st century. Major current developments. *Linguistic Typology* 11: 239–51.
Bickel, B. 2008. A refined sampling procedure for genealogical control. *Language Typology and Universals* 61: 221–33.
Bickel, B. 2010. Grammatical relations typology. In Song (ed.), 399–444.
Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N., Rai, I., Rai, M., Rai, N. & Stoll, S. 2007. Free prefix ordering in Chintang. *Language* 83(1): 43–73.
Bickel, B. & Nichols, J. 2007. Inflectional morphology. In Shopen (ed.), Vol. III: 169–240.
Bickel, B. & Nichols, J. 2011a. Exponence of selected inflectional formatives. In Dryer & Haspelmath (eds), Ch. 21. <http://wals.info/chapter/21> (11 May 2011).
Bickel, B. & Nichols, J. 2011b. Fusion of selected inflectional formatives. In Dryer & Haspelmath (eds), Ch. 20. <http://wals.info/chapter/20> (11 May 2011).
Bickerton, D. 1980[1974]. Creolization, linguistic universals, natural semantax and the brain. In *Issues in English Creoles,* R. Day (ed.), 1–18. Heidelberg: Groos.
Billings, L. 1995. Approximation in Russian and the Single-Word Constraint. PhD dissertation, Princeton University.
Bisang, W. 1999. Classifiers in East and Southeast Asian languages. Counting and beyond. In *Numeral Types and Changes Worldwide,* J. Gvozdanović (ed.), 113–86. Berlin: Mouton de Gruyter.
Bisang, W. 2009. Serial verb constructions. *Language and Linguistics Compass*. 3(3): 792–814.
Blake, B. 2001. *Case,* 2nd edn. Cambridge: CUP.
Blench, R. 2006. The Fyem language of Central Nigeria and its affinities. <http://www.rogerblench.info/RBOP.htm>
Blevins, J. 1999. Untangling Leti infixation. *Oceanic Linguistics* 38(2): 383–403.
Blevins, J. 2001a. *Nhanda. An Aboriginal Language of Western Australia.* Honolulu HI: University of Hawai'i Press.
Blevins, J. 2001b. The syllable in phonological theory. In *Phonology. Critical Concepts,* C. W. Kreidler (ed.), 75–120. London: Routledge.
Blevins, J. 2003. One case of contrast evolution in the Yurok vowel system. *International Journal of American Linguistics* 69(2): 135–50.
Blevins, J. 2004. *Evolutionary Phonology. The Emergence of Sound Patterns.* Cambridge: CUP.
Blevins, J. 2007. The importance of typology in explaining recurrent sound patterns. *Linguistic Typology* 11: 107–13.
Blust, R. 1995. Notes on Berawan consonant gemination. *Oceanic Linguistics* 34(1): 123–38.
Bodomo, A. 1997. *The Structure of Dagaare.* Stanford CA: CSLI.
Booij, G. 2005. *The Grammar of Words. An Introduction to Linguistic Morphology.* Oxford: OUP.
Booij, G. E., Lehmann, C., Mugdan, J., Skopeteas, S. & Kesselheim, W. (eds). 2000. *An International Handbook on Inflection and Word-Formation.* Berlin: Walter de Gruyter.
Bollée, A. 2013. Réunion Creole. In Michaelis et al. (eds), Vol. II.
Bowe, H. 1990. *Categories, Constituents and Constituent Order in Pitjantjatjara. An Aboriginal Language of Australia.* London: Routledge.

Branson, J., Miller, D. & Marsaja, I. 1996. Everyone here speaks sign language, too. A deaf village in Bali, Indonesia. In *Multicultural Aspects of Sociolinguistics in Deaf Communities,* C. Lucas (ed.), 39–57. Washington DC: Gallaudet University Press.

Brentari, D. 1993. Establishing a sonority hierarchy in American Sign Language. The use of simultaneous structure in phonology. *Phonology* 10: 281–306.

Brenzinger, I. 1973. Notes on the phonetics of Latvian. *Journal of the International Phonetic Association* 3(1): 36–39.

Brenzinger, M. (ed.). 2008. *Language Diversity Endangered.* Berlin: Mouton de Gruyter.

Bricker, V. 1978. Antipassive constructions in Yucatec Maya. In *Papers in Mayan Linguistics,* N. C. England (ed.), 3–24. Columbia MU: Dept. of Anthropology, University of Missouri-Columbia.

Brown, C. 1983. Where do cardinal direction terms come from? *Anthropological Linguistics* 25(2): 121–61.

Brown, C. 1989. Naming the days of the week. A cross-linguistic study of lexical acculturation. *Current Anthropology* 30(4): 536–550.

Brown, K. (ed.). 2006. *Encyclopedia of Language and Linguistics*, 2nd edn. Oxford: Elsevier.

Brown, P. & Levinson, S. 1987 [1978]. *Politeness. Some Universals in Language Usage,* 19th edn. Cambridge: CUP.

Brunelli, M. 2006. The Grammar of Italian Sign Language. With a Study about Its Restrictive Relative Clauses. Honours thesis, Ca' Foscari University, Venice.

Burquest, D. 1986. The pronoun system of some Chadic languages. In Wiesemann (ed.), 71–101.

Burridge, K. 1998. From modal auxiliary to lexical verb. The curious case of Pennsylvania German *wotte*. In *Historical Linguistics 1995. Selected Papers from the 12th ICHL,* Vol. 2: *Germanic Linguistics,* R. M. Hogg & L. Bergen (eds), 19–33. Amsterdam: John Benjamins.

Burridge, K. 2006. Language contact and convergence in Pennsylvania German. In Aikhenvald & Dixon (eds), (a) 179–200.

Butskhrikidze, M. 2002. *The Consonant Phonotactics of Georgian.* Utrecht: LOT.

Büttner, T. 1983. *Las lenguas de los Andes centrales. Estudios sobre la clasificación genética, areal y tipológica.* Madrid: Magerit.

Bybee, J. 2000. Verb. In Booij et al. (eds), 794–808.

Bybee, J. 2003a. Mechanisms of change in grammaticalization. The role of frequency. In Joseph & Janda (eds), 602–23.

Bybee, J. 2003b. Cognitive processes in grammaticalization. In *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure,* M. Tomasello (ed.), 145–67. Mahwah NJ: Lawrence Erlbaum Associates.

Bybee, J. 2010. Markedness. Iconicity, economy and frequency. In Song (ed.), 131–47.

Bybee, J. L. (ed.). 2007. *Frequency of Use and the Organization of Language.* Oxford: OUP.

Bybee, J., Perkins, R. & Pagliuca, W. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World.* Chicago IL: University of Chicago Press.

Bye, P. 2007. Grade alternation in Inari Saami and abstract declarative phonology. In *Saami Linguistics,* I. Toivonen & D. C. Nelson (eds), 53–90. Amsterdam: John Benjamins.

Campbell, L. 1985. *The Pipil Language of El Salvador.* Berlin: Mouton de Gruyter.

Campbell, L. 1997. *American Indian Languages. The Historical Linguistics of Native America.* Oxford: OUP.

Campbell, L. 2004. *Historical Linguistics. An Introduction,* 2nd edn. Edinburgh: EUP.

Campbell, L. 2009. Areal linguistics. In *Concise Encyclopedia of Languages of the World,* K. Brown & S. Ogilvie (eds), 62–68. Amsterdam: Elsevier.

Campbell, L., Kaufman, T. & Smith-Stark, T. 1986. Meso-America as a linguistic area. *Language* 62(3): 530–570.
Cardoso, H. 2013. Diu Indo-Portuguese. In Michaelis et al. (eds), Vol. II.
Carlson, R. 1994. *A Grammar of Supyire.* Berlin: Mouton de Gruyter.
Carnie, A. 2002. *Syntax. A Generative Introduction,* 2nd edn. Malden MA: Blackwell.
Carson, N. 1982. Phonology and Morphosyntax of Macuxi (Carib) (Brazil). PhD dissertation, The University of Kansas.
Cecchetto, C., Geraci, C. & Zucchi, S. 2006. Strategies of relativization in Italian Sign Language. *Natural Language and Linguistic Theory* 24: 945–75.
Chambers, J. & Trudgill, P. 1998. *Dialectology,* 2nd edn. Cambridge: CUP.
Chamoreau, C. 1998. Le système verbal du Phurhépecha. In *Systèmes verbaux,* F. Bentolila (ed.), 55–69. Louvain-la-Neuve: Peeters.
Chamoreau, C. 2007. Grammatical borrowing in Purepecha. In Matras & Sakel (eds), 465–481.
Chang, L. 1998. Thao reduplication. *Oceanic Linguistics* 37(2): 277–297.
Chao, Y. 1970. The Changchow Dialect. *Journal of the American Oriental Society* 90(1): 45–56.
Chappell, H. 2008. Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages. *Linguistic Typology* 12: 45–98.
Charney, J. 1993. *A Grammar of Comanche.* Lincoln NB: University of Nebraska Press.
Cheetham, B. 1978. Counting and number in Huli. *Papua New Guinea Journal of Education* 14: 16–27.
Chelliah, S. 1997. *A Grammar of Meithei.* Berlin: Mouton de Gruyter.
Childs, G. 1995. *A Grammar of Kisi. A Southern Atlantic language.* Berlin: Mouton de Gruyter.
Chung, S. & Timberlake, A. 1985. Tense, aspect, and mood. In *Language Typology and Syntactic Description*. Vol III: *Grammatical Categories and the Lexicon*, T. Shopen (ed.), 202–258. Cambridge: CUP
Claudi, U. 1994. Word order change as category change. The Mande case. In *Perspectives on Grammaticalization,* W. Pagliuca (ed.), 191–231. Amsterdam: John Benjamins.
Cole, P., Hara, Y. & Thai Yap, N. 2008. Auxiliary fronting in Peranakan Javanese. *Journal of Linguistics* 44: 1–43.
Comrie, B. 1976. *Aspect.* Cambridge: CUP.
Comrie, B. 1981a. Ergativity and grammatical relations in Kalaw Lagaw Ya (Saibai dialect). *Australian Journal of Linguistics* 1: 1–42.
Comrie, B. 1981b. *Language Universals and Linguistic Typology.* 1st edn. Oxford: Blackwell.
Comrie, B. 1981c. *Languages of the Soviet Union.* Cambridge: CUP.
Comrie, B. 1985. *Tense.* Cambridge: CUP.
Comrie, B. 1989. *Language Universals and Linguistic Typology*. *Syntax and Morphology,* 2nd edn. Oxford: Blackwell.
Comrie, B. 2011a. Alignment of case marking of full noun phrases. In Dryer & Haspelmath (eds), Feature 98A. <http://wals.info/feature/98A> (11 May 2011).
Comrie, B. 2011b. Alignment of case marking of pronouns. In Dryer & Haspelmath (eds), Feature 99A. <http://wals.info/feature/99A> (11 May 2011).
Comrie, B. 2011c. Numeral bases. In Dryer & Haspelmath (eds), Ch. 131. <http://wals.info/chapter/131> (11 May 2011).
Comrie, B. & Keenan, E. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8(1): 63–99.
Comrie, B. & Kuteva, T. 2011a. Relativization on obliques. In Dryer & Haspelmath (eds), Ch. 123. <http://wals.info/chapter/123> (11 May 2011).

Comrie, B. & Kuteva, T. 2011b. Relativization on subjects. In Dryer & Haspelmath (eds), Ch. 122. <http://wals.info/chapter/122> (11 May 2011).
Comrie, B. & Kuteva, T. 2011c. Relativization strategies. In Dryer & Haspelmath (eds), Supplement 8. <http://wals.info/supplement/8> (11 May 2011).
Comrie, B. & Polinsky, M. 1998. The great Daghestan case coax. In *Case, Typology and Grammar. In Honor of Barry J. Blake,* A. Siewierska & J. J. Song (eds), 95–114. Amsterdam: John Benjamins.
Corbett, G. 1991. *Gender.* Cambridge: CUP.
Corbett, G. 1993. The head of Russian numeral expressions. In *Heads in Grammatical Theory,* G. G. Corbett, N. M. Fraser & S. McGlashan (eds), 11–35. Cambridge: CUP.
Corbett, G. 2000. *Number.* Cambridge: CUP.
Corbett, G. 2006. *Agreement.* Cambridge: CUP.
Corbett, G. 2007. Gender and noun classes. In Shopen (ed.), Vol. III: 241–279.
Corbett, G. 2011a. Number of genders. In Dryer & Haspelmath (eds), Ch. 30. <http://wals.info/chapter/30> (11 May 2011).
Corbett, G. 2011b. Sex-based and non-sex-based gender systems. In Dryer & Haspelmath (eds), Ch. 31. <http://wals.info/chapter/31> (11 May 2011).
Corina, D. & Sandler, W. 1993. On the nature of phonological structure in sign language. *Phonology* 10: 165–207.
Cornyn, W. 1944. Outline of Burmese Grammar. *Language* 20(4): 5–34.
Corston-Oliver, S. 2002. Roviana. In Lynch et al. (eds), 467–97.
Cristofaro, S. 2005. *Subordination,* reprinted with corrections. Oxford: OUP.
Cristofaro, S. 2010. Language universals and linguistic knowledge. In Song (ed.), 227–249.
Croft, W. 1990. *Typology and Universals,* 1st edn. Cambridge: CUP.
Croft, W. 1991. The evolution of negation. *Journal of Linguistics* 27: 1–27.
Croft, W. 2000. *Explaining Language Change. An Evolutionary Approach.* Harlow: Longman.
Croft, W. 2002. *Radical Construction Grammar. Syntactic Theory in Typological Perspective,* reprint, with corrections. Oxford: OUP.
Croft, W. 2003. *Typology and Universals,* 2nd edn. Cambridge: CUP.
Crowley, T. 1998. *An Erromangan (Sye) Grammar.* Honolulu HI: University of Hawai'i Press.
Crowley, T. 2002. Gela. In Lynch et al. (eds), 525–537.
Crowley, T. 2007. *Field Linguistics. A Beginner's Guide.* Oxford: OUP.
Cruz-Ferreira, M. 2010. *Multilinguals are …?* London: Battlebridge.
Culicover, P. & Jackendoff, R. 2005. *Simpler Syntax.* Oxford: OUP.
Curnow, T. 2001. What language features can be borrowed? In Aikhenvald & Dixon (eds), 412–436.
Cysouw, M. 2005. A typology of honorific uses of clusivity. In *Clusivity. Typology and Case Studies of Inclusive-Exclusive Distinction,* E. Filimonova (ed.), 213–230. Amsterdam: John Benjamins.
Cysouw, M. 2011. Inclusive/exclusive forms in independent pronouns. In Dryer & Haspelmath (eds), Feature 39A. <http://wals.info/feature/39A> (11 May 2011).
Dahl, Ö. 1985. *Tense and Aspect Systems.* Oxford: Blackwell.
Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity.* Amsterdam: John Benjamins.
Dahl, Ö. & Velupillai, V. 2011a. Perfective/imperfective aspect. In Dryer & Haspelmath (eds), Ch. 65. <http://wals.info/chapter/65> (11 May 2011).
Dahl, Ö. & Velupillai, V. 2011b. The future tense. In Dryer & Haspelmath (eds), Ch. 67. <http://wals.info/chapter/67> (11 May 2011).
Dahl, Ö. & Velupillai, V. 2011c. The past tense. In Dryer & Haspelmath (eds), Ch. 66. <http://wals.info/chapter/66> (11 May 2011).

Dahl, Ö. & Velupillai, V. 2011d. The perfect. In Dryer & Haspelmath (eds), Ch. 58. <http://wals.info/chapter/68> (11 May 2011).
Dahlstrom, A. 1986. Plains Cree Morphosyntax (Algonquian). PhD dissertation, University of California, Berkeley.
Dalrymple, M. & Nikolaeva, I. 2011. *Objects and Information Structure.* Cambridge: CUP.
Daniel, M. 2011. Plurality in independent personal pronouns. In Dryer & Haspelmath (eds), Ch. 35. <http://wals.info/chapter/35> (11 May 2011).
Daniel, M. & Moravcsik, E. 2011. The associative plural. In Dryer & Haspelmath (eds), Ch. 36. <http://wals.info/chapter/36> (11 May 2011).
Das Gupta, K. 1971. *An Introduction to the Nocte Language.* Shillong: North East Frontier Agency.
Davidson, M. 2002. Studies in Wakashan (Nootkan) Grammar. PhD dissertation, University of New York at Buffalo.
Davies, W. 1986. *Choctaw Verb Agreement and Universal Grammar.* Dordrecht: Reidel.
Dayley, J. 1981. Tzutujil Grammar. PhD dissertation, University of California, Berkeley.
de Haan, F. 2010. Typology of tense, aspect, and modality systems. In Song (ed.), 445–64.
de Haan, F. 2011. Semantic distinctions of evidentiality. In Dryer & Haspelmath (eds), Ch. 77. <http://wals.info/chapter/77> (11 May 2011).
de Rooij, V. 2007. Grammatical borrowing in Katanga Swahili. In Matras & Sakel (eds), 123–37.
DeGraff, M. 2009. Language acquisition in creolization and, thus, language change. Some Cartesian-Uniformitarian boundary conditions. *Language and Linguistics Compass* 3(4): 888–971.
DeLancey, S. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3): 626–57.
DeLancey, S. 2006. The blue bird of ergativity. Ms. <http://celia.cnrs.fr/FichExt/Documents%20de%20travail/Ergativite/3dDelancey.pdf>.
Demolin, D. 1991. L'analyse des segments, de la syllabe et des tons dans un jeu de langage mangbetu. *Langages* 101: 30–50.
Dench, A. 1995. *Martuthunira. A Language of the Pilbara Region of Western Australia.* Canberra: Pacific Linguistics.
Derbyshire, D. 1977. Word order universals and the existence of OVS languages. *Linguistic Inquiry* 8(3): 590–599.
Derhemi, E. (ed.). 2002. The impact of government policies on territorially based ethnic or nationalist movements. *International Journal on Multicultural Societies* 4: 2.
Deuchar, M. 1987. Negative incorporation in three Sign Languages. In *Sign and School. Using Signs in Deaf Children's Development,* J. Kyle (ed.), 35–43. Clevedon: Multilingual Matters.
Dickens, P. n.y. *Ju/'hoan Grammar*. Ms.
Die Brücke zur Welt der Gehörlosen 2001. Würzburg: Bezirksverband der Gehörlosen Unterfranken e.V.
Diessel, H. 1999a. *Demonstratives. Form, Function, and Grammaticalization.* Amsterdam: John Benjamins.
Diessel, H. 1999b. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3: 1–49.
Diessel, H. 2011. Distance contrasts in demonstratives. In Dryer & Haspelmath (eds), Ch. 41. <http://wals.info/chapter/41> (11 May 2011).
Dimmendaal, G. 2001. Areal diffusion and genetic inheritance. An African perspective. In Aikhenvald & Dixon (eds), 358–392.
Dixon, R. 1972. *The Dyirbal Language of North Queensland.* Cambridge: CUP.
Dixon, R. 1979. Ergativity. *Language* 55(1): 59–138.
Dixon, R. 1982. *Where Have all the Adjectives Gone? And Other Essays in Semantics and Syntax.* Berlin: Walter de Gruyter.

Dixon, R. 1988. *A Grammar of Boumaa Fijian.* Chicago IL: University of Chicago Press.
Dixon, R. 1994. *Ergativity.* Cambridge: CUP.
Dixon, R. 2000. A typology of causatives. Form, syntax and meaning. In *Changing Valency. Case Studies in Transitivity,* R. M. W. Dixon & A. Aikhenvald (eds), 30–83. Cambridge: CUP.
Dixon, R. 2006a. Adjective classes in typological perspective. In *Adjective Classes. A Cross-Linguistic Typology,* A. Aikhenvald & R. M. W. Dixon (eds), 1–49. Oxford: OUP.
Dixon, R. 2006b. Complement clause types and complementation strategies in typological perspective. In *Complementation. A Cross-Linguistic Typology,* R. M. W. Dixon & A. Aikhenvald (eds), 1–48. Oxford: OUP.
Dixon, R. M. W. & Aikhenvald, A. (eds). 2002. *Word. A Cross-Linguistic Typology.* Cambridge: CUP.
Dixon, R. M. W. & Aikhenvald, A. (eds). 2006. *Complementation. A Cross-Linguistic Typology.* Oxford: OUP.
Doerfer, G. 1966. Zur Verwandschaft der altaischen Sprachen. *Indogermanische Forschungen* 71: 81–122.
Dol, P. 1999. A grammar of Maybrat. A Language of the Bird's Head, Irian Jaya, Indonesia. PhD dissertation, University of Leiden.
Donaldson, T. 1980. *Ngiyambaa. The Language of the Wangaaybuwan.* Cambridge: CUP.
Donohue, M. 1996. Bajau. A symmetrical Austronesian language. *Language* 72(4): 782–793.
Donohue, M. 1999. *A Grammar of Tukang Besi.* Berlin: Mouton de Gruyter.
Dooley, R. 2010. Exploring clause chaining. *SIL Electronic Work Papers,* 2010-001(February), 20 pages. <http://www.sil.org/silewp/2010/silewp2010-001.pdf>
Dorvlo, K. 2008. *A Grammar of Logba (Ikpana).* Utrecht: LOT.
Dotter, F. 2001. Gebärdensprachforschung. In Haspelmath et al. (eds), 141–53.
Dryer, M. 1989a. Large linguistic areas and language sampling. *Studies in Language* 13(2): 257–292.
Dryer, M. 1989b. Salinan word order. In *Papers from the 1988 Hokan-Penutian Languages Workshop,* S. DeLancey (ed.), 40–49. Eugene OR: Department of Linguistics, University of Oregon.
Dryer, M. 1997. Why statistical universals are better than absolute universals. *Chicago Linguistic Society,* 33(2): 123–145.
Dryer, M. 2005a. Order of adjective and noun. In Haspelmath et al. (eds), 354–357.
Dryer, M. 2005b. Order of demonstrative and noun. In Haspelmath et al. (eds), 358–361.
Dryer, M. 2005c. Order of genitive and noun. In Haspelmath et al. (eds), 350–353.
Dryer, M. 2005d. Order of numeral and noun. In Haspelmath et al. (eds). 362–365.
Dryer, M. 2007a. Clause types. In Shopen (ed.), Vol. I: 224–275.
Dryer, M. 2007b. Word order. In Shopen (ed.), Vol. I: 61–131.
Dryer, M. 2011a. Definite articles. In Dryer & Haspelmath (eds), Ch. 37. <http://wals.info/chapter/37> (11 May 2011).
Dryer, M. 2011b. Determining dominant word order. In Dryer & Haspelmath (eds), Supplement 6. <http://wals.info/supplement/6> (11 May 2011).
Dryer, M. 2011c. Expression of pronominal subjects. In Dryer & Haspelmath (eds), Ch. 101. <http://wals.info/chapter/101> (11 May 2011).
Dryer, M. 2011d. Genealogical language list. In Dryer & Haspelmath (eds), Supplement 4. <http://wals.info/supplement/4> (11 May 2011).
Dryer, M. 2011e. Indefinite articles. In Dryer & Haspelmath (eds), Ch. 38. <http://wals.info/chapter/38> (11 May 2011).
Dryer, M. 2011f. Languages with two dominant orders of subject, object and verb. In Dryer & Haspelmath (eds), Feature 81B. <http://wals.info/feature/81B> (11 May 2011).

Dryer, M. 2011g. Negative morphemes. In Dryer & Haspelmath (eds), Ch. 112. <http://wals.info/chapter/112> (11 May 2011).
Dryer, M. 2011h. Order of adjective and noun. In Dryer & Haspelmath (eds), Ch. 87. <http://wals.info/chapter/87> (11 May 2011).
Dryer, M. 2011i. Order of adposition and noun phrase. In Dryer & Haspelmath (eds), Ch. 85. <http://wals.info/chapter/85> (11 May 2011).
Dryer, M. 2011j. Order of adverbial subordinator and clause. In Dryer & Haspelmath (eds), Ch. 94. <http://wals.info/chapter/94> (11 May 2011).
Dryer, M. 2011k. Order of demonstrative and noun. In Dryer & Haspelmath (eds), Ch. 88. <http://wals.info/chapter/88> (11 May 2011).
Dryer, M. 2011l. Order of genitive and noun. In Dryer & Haspelmath (eds), Ch. 86. <http://wals.info/chapter/86> (11 May 2011).
Dryer, M. 2011m. Order of negative morpheme and verb. In Dryer & Haspelmath (eds), Ch. 143. <http://wals.info/chapter/143> (11 May 2011).
Dryer, M. 2011n. Order of numeral and noun. In Dryer & Haspelmath (eds), Ch. 89. <http://wals.info/chapter/89> (11 May 2011).
Dryer, M. 2011o. Order of object and verb. In Dryer & Haspelmath (eds), Ch 83. <http://wals.info/chapter/83> (11 May 2011).
Dryer, M. 2011p. Order of relative clause and noun. In Dryer & Haspelmath (eds), Feature 90A. <http://wals.info/feature/90A> (11 May 2011).
Dryer, M. 2011q. Order of subject and verb. In Dryer & Haspelmath (eds), Ch. 82. <http://wals.info/chapter/82> (11 May 2011).
Dryer, M. 2011r. Order of subject, object, and verb. In Dryer & Haspelmath (eds), Feature 81A. <http://wals.info/feature/81A> (11 May 2011).
Dryer, M. 2011s. Polar questions. In Dryer & Haspelmath (eds), Ch. 116. <http://wals.info/chapter/116> (11 May 2011).
Dryer, M. 2011t. Position of case affixes. In Dryer & Haspelmath (eds), Ch. 51, <http://wals.info/chapter/51> (11 May 2011).
Dryer, M. 2011u. Position of interrogative phrases in content questions. In Dryer & Haspelmath (eds), Ch. 93. <http://wals.info/chapter/93> (11 May 2011).
Dryer, M. 2011v. Position of polar question particles. In Dryer & Haspelmath (eds), Ch. 92. <http://wals.info/chapter/92> (11 May 2011).
Dryer, M. 2011w. Prefixing versus suffixing in inflectional morphology. In Dryer & Haspelmath (eds), Ch. 26. <http://wals.info/chapter/26> (11 May 2011).
Dryer, M. 2011x. Relationship between the order of object and verb and the order of adjective and noun. In Dryer & Haspelmath (eds), Ch 97. <http://wals.info/chapter/97> (11 May 2011).
Dryer, M. 2011y. Relationship between the order of object and verb and the order of adposition and noun phrase. In Dryer & Haspelmath (eds), Ch. 95. <http://wals.info/chapter/95> (11 May 2011).
Dryer, M. 2011z. Relationship between the order of the object and verb and the order of the relative clause and noun. In Dryer & Haspelmath (eds), Ch. 96. <http://wals.info/chapter/96> (11 May 2011).
Dryer, M. & Haspelmath, M. (eds). 2011. *The World Atlas of Language Structures Online.* München: MPDL.
Du Feu, V. 1996. *Rapanui.* London: Routledge.
Eastman, C. 1992. Swahili interjections. Blurring language-use/gesture-use boundaries. *Journal of Pragmatics* 18: 273–287.

Eberhard, M. 2009. *Mamaindê Grammar. A Northern Nambikwara Language and Its Cultural Context.* Utrecht: LOT.

Echegoyen, A. 1979. *Luces contemporáneas del otomí. Gramática del otomí de la sierra.* Mexico: Instituto Lingüístico de Verano.

Ehrhart, S. & Halpap, M. 2013. Tayo. In Michaelis et al. (eds), Vol. II.

Einaudi, P. 1974. A Grammar of Biloxi. PhD dissertation, University of Colorado at Boulder.

Elert, C.-C. 1989. *Allmän och svensk fonetik,* 6th edn. Stockholm: Norstedts.

Emeneau, M. 1956. India as a linguistic area. *Language* 32(1): 3–16.

Emmorey, K. 2002. *Language, Cognition, and the Brain. Insights from Sign Langauge Research.* Mahwah NJ: Lawrence Erlbaum Associates.

Emmorey, K., Borinstein, H. & Thompson, R. 2005. Bimodal bilingualism. Code-switching between spoken Engish and American Sign Language. In *ISB4. Proceedings of the 4th International Symposium on Bilingualism,* J. Cohen, K. T. McAlister, K. Rolstand & J. McSwan (eds), 663–73. Somerville MA: Cascadilla Press.

Enfield, N. 2001. On genetic and areal linguistics in mainland South-East Asua. Parallel polyfunctionality of 'aquire'. In Aikhenvald & Dixon (eds), 255–90.

Enfield, N. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34: 181–206.

Enger, H.-O. 2002. The story of Scandinavian *-s(t)* retold. Grammaticalising a clitic to a derivational affix. *Folia Linguistica Historica* 23(1-2): 79–106.

Engstrand, O. 1990. Illustrations of the IPA. Swedish. *Journal of the International Phonetic Association* 20(1): 42–44.

Epps, P. 2007. Grammatical borrowing in Hup. In Matras & Sakel (eds), 551–66.

Escure, G. 2013. Belizean Creole. In Michaelis et al. (eds), Vol. I.

Essien, O. 1982. The so-called reflexive pronouns and reflexivization in Ibibio. *Studies in African Linguistics* 13(2): 93–108.

Evans, N. 1995. *A Grammar of Kayardild. With Historical-Comparative Notes on Tangkic.* Berlin: Mouton de Gruyter.

Evans, N. 2000. Word classes in the world's languages. In Booij et al. (eds), 708–32.

Everett, D. 1982. Phonetic rarities in Pirahã. *Journal of the International Phonetic Association* 12(2): 94–96.

Everett, D. 1998. Wari (Amazonian). In Spencer & Zwicky (eds), 690–706.

Everett, D. 2005. Cultural constraints on grammar and cognition in Pirahã. Another look at the design features of human language. *Current Anthropology* 2005: 621–46.

Everett, D. & Kern, B. 1997. *Wari'. The Pacaas Novos Language of Western Brazil.* London: Routledge.

Fabre, A. 2009. Diccionario etnolingüístico y guía bibliográfica de los pueblos indígenas sudamericanos. <http://butler.cc.tut.fi/~fabre/BookInternetVersio/Alkusivu.html> (Accessed 19 August 2011).

Faraclas, N. 2013. Nigerian Pidgin. In Michaelis et al. (eds), Vol. I.

Farquharson, J. 2013. Jamaican. In Michaelis et al. (eds), Vol. I.

Farrell, P. 2005. *Grammatical Relations.* Oxford: OUP.

Fattier, D. 2013. Haitian Creole. In Michaelis et al. (eds), Vol. II.

Férère, G. 1983. Nasalized vowels and semiconsonants in Haitian Creole. *Journal of the International Phonetic Association* 13(2): 76–81.

Ferguson, C. 1976. The Ethiopian language area. In *Language in Ethiopia,* M. L. Bender, J. D. Bowen, R. L. Cooper & C. A. Ferguson (eds), 63–76. Oxford: OUP.

Fernández-Viader, M. & Fuentes, M. 2008. The systems of numerals in Catalan Sign language (LSC) and Spanish Sign Language (LSE): A comparative study. In Müller de Quadros (ed.), 96–106.
Fischer, S. 1973. Two processes of reduplication in the American Sign Language. *Foundations of Language* 9(4): 469–480.
Fischer, S. 2007. Grammatical borrowing in Rapanui. In Matras & Sakel (eds), 387–402.
Fleck, D. 2007. Evidentiality and double tense in Matses. *Language* 83(3): 589–614.
Foley, W. 2007. A typology of information packaging in the clause. In Shopen (ed.), Vol. I: 362–446.
Foley, W. 2013. Yimas-Arafundi Pidgin. In Michaelis et al. (eds), Vol. III.
Foley, W. & Van Valin, R. 1984. *Functional Syntax and Universal Grammar.* Cambridge: CUP.
Fortescue, M. 1984. *West Greenlandic.* London: Croom Helm.
Frajzyngier, Z. 1987. From verb to anaphora. *Lingua* 72: 155–68.
Frajzyngier, Z. & Shay, E. 2002. *A Grammar of Hdi.* Berlin: Mouton de Gruyter.
François, A. 2005. Language profile. A typological overview of Mwotlap, an Oceanic language of Vanuatu. *Linguistic Typology* 9: 115–46.
French, K. 1988. *Insights into Tagalog. Reduplication, Infixation, and Stress from Nonlinear Phonology.* Dallas TX: SIL.
Fridman-Mintz, B. 2005. Tense and Aspect Inflections in Mexican Sign Language Verbs. PhD dissertation, Georgetown University.
Gabas, N., Jr 1999. A Grammar of Karo, Tupi (Brazil). PhD dissertation, University of California at Santa Barbara.
Gaby, A. 2006. A Grammar of Kuuk Thaayorre. PhD dissertation, The University of Melbourne.
Geis, M. 2006. *Speech Acts and Conversational Interaction.* Cambridge: CUP.
Georgopoulos, C. 1991. *Syntactic Variables. Resumptive Pronouns and A' Binding in Palauan.* Dordrecht: Kluwer.
Gerdts, D. 1998. Incorporation. In Spencer & Zwicky (eds), 84–100.
Gerner, M. 2004. On a partial, strictly word-order based definition of grammatical relations in Liangshan Nuosu. *Linguistics* 42(1): 109–54.
Ghaniabadi, S., Ghomeshi, J. & Sadat-Tehrani, N. 2006. Reduplication in Persian. A morphological doubling approach. In *Proceedings of the 2006 Canadian Linguistics Association Annual Conference,* C. Gurski & M. Radisic (eds). 15 pages. <http://ling.uwo.ca/publications/CLA2006/Ghaniabadi_etal.pdf>
Gil, D. 2001. Escaping Eurocentrism. Fieldwork as a process of unlearning. In Newman & Ratliff (eds), 102–132.
Gil, D. 2011. Numeral classifiers. In Dryer & Haspelmath (eds), Ch. 55. <http://wals.info/chapter/55> (11 May 2011).
Gippert, J., Himmelmann, N. & Mosel, U. (eds). 2006. *Essentials of Language Documentation.* Berlin: Mouton de Gruyter.
Givón, T. 1984. Direct object and dative shifting. Semantic and pragmatic case. In *Objects. Towards a Theory of Grammatical Relations,* F. Plank (ed.), 151–170. New York NY: Academic Press.
Givón, T. 2001a. *Syntax. An Introduction*. Vol. I, 2nd edn. Amsterdam: John Benjamins.
Givón, T. 2001b. *Syntax. An Introduction*. Vol. II. 2nd edn. Amsterdam: John Benjamins.
Glass, A. & Hackett, D. 1970. *Pitjantjatjara Grammar. A Tagmemic View of Ngaanyatjara (Warburton Ranges) Dialect.* Canberra: Australian Institute of Aboriginal Studies.
Glinert, L. 1989. *The Grammar of Modern Hebrew.* Cambridge: CUP.
Goedemans, R. & van der Hulst, H. 2011a. Fixed stress locations. In Dryer & Haspelmath (eds), Ch. 14. <http://wals.info/chapter/14> (11 May 2011).

Goedemans, R. & van der Hulst, H. 2011b. Weight-sensitive stress. In Dryer & Haspelmath (eds), Ch. 15. <http://wals.info/chapter/15> (11 May 2011).
Goswami, G. & Tamuli, J. 2007. Asamiya. In *The Indo-Aryan Languages,* G. Cardona & D. Jain (eds), 429–84. London: Routledge.
Goudswaard, N. 2005. *The Begak (Ida'an) Language of Sabah.* Utrecht: LOT.
Graffi, G. 2010. The pioneers of linguistic typology. From Gabelentz to Greenberg. In Song (ed.), 25–42.
Grant, A. 1999. Mixed languages. A conspectus. Ms.
Grant, A. 2002. Fabric, pattern, shift and diffusion. What change in Oregon Penutian languages can tell historical linguists. In *Proceedings of the Meeting of the Hokan-Penutian Workshop, June 17–18, 2000, U. of California at Berkeley. Report 11, Survey of California and Other Indian Languages,* L. Buszard-Welcher (ed.), 33–56. Berkeley CA: University of California at Berkeley.
Grant, A. 2003. Reduplication in Mindanao Chabacano. In *Twice as Meaningful. Reduplication in Pidgins, Creoles and Other Contact Languages,* S. Kouwenberg (ed.), 203–10. London: Battlebridge.
Grant, A. 2013. Chinuk Wawa. In Michaelis et al. (eds), Vol. III.
Greenberg, J. 1960[1954]. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26: 178–194.
Greenberg, J. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language,* J. H. Greenberg (ed.), 58–90. Cambridge MA: The MIT Press.
Greenberg, J. 1978. Diachrony, synchrony, and language universals. In *Universals of Human Language,* Vol. I, J. H. Greenberg, C. A. Ferguson & E. Moravcsik (eds), 61–92. Stanford CA: Stanford University Press.
Greenberg, J. 1990. Some universals of grammar with particular reference to the order of meaningful elements. In *On Language. Selected Writings of Joseph H. Greenberg,* K. Denning & S. Kemmer (eds), 40–70. Stanford CA: Stanford University Press.
Grenand, F. 1980. *La langue wayãpi (Guyane française). Phonologie et grammaire.* Paris: Société d'études linguistiques et anthropologiques de France.
Gries, S. 2009. *Statistics for Linguistics with R. A Practical Introduction.* Berlin: Mouton de Gruyter.
Grinevald, C. n.y. *A Grammar of Rama*. Ms.
Grønnum, N. 1998. Illustrations of the IPA. Danish. *Journal of the International Phonetic Association* 28(1-2): 99–105.
Gulya, J. 1966. *Eastern Ostyak Chrestomathy.* Bloomington IN: Indiana University Press.
Gumperz, J. 1982. *Discourse Strategies.* Cambridge: CUP.
Hagège, C. 2008. Towards a typology of interrogative verbs. *Linguistic Typology* 12: 1–44.
Haiman, J. 1980a. *Hua. A Papuan Language of the Eastern Highlands of New Guinea.* Amsterdam: John Benjamins.
Haiman, J. 1980b. The iconicity of grammar. Isomorphism and motivation. *Language* 56(3): 515–40.
Haiman, J. 1983. Iconic and economic motivation. *Language* 59(4): 781–819.
Haiman, J. 2010. Competing motivations. In Song (ed.), 148–65.
Haiman, J. & Thompson, S. A. (eds). 1988. *Clause Combining in Grammar and Discourse.* Amsterdam: John Benjamins.
Hajek, J. 2011. Vowel nasalization. In Dryer & Haspelmath (eds), Ch. 10. <http://wals.info/chapter/10> (11 May 2011).
Hale, K. 1992. Language endangerment and the human value of linguistic diversity. *Language* 68(1): 35–42.

Hamel, P. 1993. Serial verbs in Loniu and an evolving preposition. *Oceanic Linguistics* 32(1): 111–132.
Hammarström, H. 2009. Sampling and genealogical coverage in WALS. *Linguistic Typology* 13: 105–19.
Hammarström, H. 2010. Rarities in numeral systems. In *Rara & Rarissima. Collecting and Interpreting Unusual Characteristics of Human Languages,* J. Wohlgemuth & M. Cysouw (eds), 7–55. Berlin: Mouton de Gruyter.
Harley, M. & Blench, R. Forthcoming. Unusual sounds in Nigerian languages. Ms.
Harlow, R. 2007. *Māori. A Linguistic Introduction.* Cambridge: CUP.
Harms, P. 1994. *Epena Pedee Syntax.* Arlington TX: SIL.
Harris, A. 2000. Where in the word is the Udi clitic? *Language* 76(3): 593–616.
Harris, A. & Campbell, L. 1995. *Historical Syntax in Cross-Linguistic Perspective.* Cambridge: CUP.
Harrison, K. 2007. *When Languages Die. The Extinction of the World's Languages and the Erosion of Human Knowledge.* Oxford: OUP.
Harvey, M. 2002. *A Grammar of Gaagudju.* Berlin: Mouton de Gruyter.
Harvey, M. n.y. Warray Grammar. Ms.
Haspelmath, M. 1993. *A Grammar of Lezgian.* Berlin: Mouton de Gruyter.
Haspelmath, M. 1995. The converb as a cross-linguistically valid category. In *Converbs in Cross-Linguistic Perspective. Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds,* M. Haspelmath & E. König (eds), 1–56. Berlin: Mouton de Gruyter.
Haspelmath, M. 1998. How young is Standard Average European? *Language Sciences* 20(3): 271–287.
Haspelmath, M. 2001. Word classes and parts of speech. In *International Encyclopedia of the Social and Behavioral Sciences,* P. B. Baltes & N. J. Smelser (eds), 16538–16545. Oxford: Pergamon.
Haspelmath, M. 2002. *Understanding Morphology.* London: Arnold.
Haspelmath, M. 2004. On directionality in language change with particular reference to grammaticalization. In *Up and Down the Cline. The Nature of Grammaticalization,* O. Fischer, M. Norde & H. Perridon (eds), 14–44. Amsterdam: John Benjamins.
Haspelmath, M. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1): 25–70.
Haspelmath, M. 2007. Coordination. In Shopen (ed.), Vol. II: 1–51.
Haspelmath, M. 2011a. Ditransitive constructions. The verb 'give'. In Dryer & Haspelmath (eds), Ch. 105. <http://wals.info/chapter/105> (11 May 2011).
Haspelmath, M. 2011b. Indefinite pronouns. In Dryer & Haspelmath (eds), Ch. 46. <http://wals.info/chapter/46> (11 May 2011).
Haspelmath, M. 2011c. Nominal and verbal conjunction. In Dryer & Haspelmath (eds), Ch. 64. <http://wals.info/chapter/64> (11 May 2011).
Haspelmath, M. 2011d. The occurrence of nominal plurality. In Dryer & Haspelmath (eds), Ch. 34. <http://wals.info/chapter/34> (11 May 2011).
Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B. (eds). 2005. *The World Atlas of Language Structures.* Oxford: OUP.
Haspelmath, M., König, E., Oesterreicher, W. & Raible, W. (eds). 2001. *Language Typology and Language Universals. An International Handbook.* Berlin: de Gruyter.
Haspelmath, M. & Tadmor, U. (eds). 2009. *Loanwords in the World's Languages. A Comparative Handbook.* Berlin: Mouton de Gruyter.
Haviland, J. 1979a. Guugu Yimidhirr. Sketch grammar. In *Handbook of Australian Languages* 1, R. M. W. Dixon & B. J. Blake (eds), 27–180. Canberra: ANU Press.
Haviland, J. 1979b. Guugu Yimidhirr brother-in-law language. *Language in Society* 8(3): 365–93.

Hawkins, J. 2001. The role of processing principles in explaining language universals. In Haspelmath et al. (eds), 360–368.
Hawkins, J. 2010. Processing efficiency and complexity in typological patterns. In Song (ed.), 206–226.
Hayward, R. 1990. Notes on the Aari language. In *Omotic Language Studies,* R. J. Hayward (ed.), 425–93. London: School of Oriental and African Studies, University of London.
Heath, J. 1984. *A Functional Grammar of Nunggubuyu.* Canberra: Australian Institute of Aboriginal Studies.
Heath, J. 1999. *A Grammar of Koyra Chiini. The Songhay of Timbuktu.* Berlin: Mouton de Gruyter.
Heath, T. 2003. Makaa (A83). In *The Bantu Languages,* D. Nurse & G. Philippson (eds), 335–348. London: Routledge.
Heine, B. 2003. Grammaticalization. In Joseph & Janda (eds), 575–601.
Heine, B., Claudi, U. & Hünnemeyer, F. 1991. *Grammaticalization. A Conceptual Framework.* Chicago IL: University of Chicago Press.
Heine, B. & Kuteva, T. 2002. *World Lexicon of Grammaticalization.* Cambridge: CUP.
Heine, B. & Kuteva, T. 2007. *The Genesis of Grammar. A Reconstruction.* Oxford: OUP.
Heine, B. & Reh, M. 1984. *Grammaticalization and Reanalysis in African Languages.* Hamburg: Helmut Buske.
Hekking, E. & Bakker, D. 2007. The case of Otomi. A contribution to Grammatical Borrowing in Cross-Linguistic Perspective. In Matras & Sakel (eds), 435–464.
Hellenthal, A.-C. 2007. Modality properties of sentence type markers in Sheko. *Leiden Papers in Linguistics* 4(2): 17–32.
Hellquist, E. 1980. *Svensk etymologisk ordbok,* 3rd edn. Malmö: Gleerups.
Helmbrecht, J. 2011. Politeness distinctions in pronouns. In Dryer & Haspelmath (eds), Ch. 45. <http://wals.info/chapter/45> (11 May 2011).
Hendriks, B. 2008. *Jordanian Sign Language. Aspects of Grammar from a Cross-Linguistic Perspective.* Utrecht: LOT.
Hengeveld, K., Rijkhoff, J. & Siewierska, A. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40: 527–570.
Hengeveld, K. & van Lier, E. 2010. The implicational map of parts-of-speech. *Linguistic Discovery* 8(1): 129–156.
Hepner, M. 2006. *Bargam Grammar Sketch.* Arlington TX: SIL.
Herder, J. 1772. *Abhandlung über den Ursprung der Sprache.* Berlin: Christian Frierich Voss.
Hildebrandt, K. 2007. Grammatical borrowing in Manange. In Matras & Sakel (eds), 283–300.
Hill, J. 2005. *A Grammar of Cupeño.* Berkeley CA: University of California Press.
Himmelmann, N. & Ladd, D. 2008. Prosodic description. An introduction for fieldworkers. *Language Documentation and Conservation* 2(2): 244–274.
Hinton, B. 1991. Aspects of Tugun Phonology and Syntax. MA thesis, University of Texas at Arlington.
Hionides, H. 2002. *Collins Pocket Greek Dictionary.* Glasgow: Harper Collins.
Hock, H. 2003. Analogical change. In Joseph & Janda (eds), 441–461.
Holm, J. 1989. *Pidgins and Creoles,* Vol. II: *Reference Survey.* Cambridge: CUP.
Holm, J. 2000. *An Introduction to Pidgin and Creoles.* Cambridge: CUP.
Holm, J. & Patrick, P. (eds). 2007. *Comparative Creole Syntax. Parallel Outlines of 18 Creole Grammars.* London: Battlebridge.
Holst, J. 2001. *Lettische Grammatik.* Hamburg: Helmut Buske.
Hopper, P. & Thompson, S. 1993. Language universals, discourse pragmatics, and semantics. *Language Sciences*. 15(4): 357–376.

Hopper, P. & Traugott, E. 2003. *Grammaticalization.* 2nd edn. Cambridge: CUP.
Hualde, J., Lujanbio, O. & Torreira, F. 2008. Lexical tone and stress in Goizueta Basque. *Journal of the International Phonetic Association* 38(1): 1–24.
Hualde, J. & Ortiz de Urbina, J. 2003. *A Grammar of Basque.* Berlin: Mouton de Gruyter.
Huber, M. 1999. *Ghanaian Pidgin English in its West African Context. A Sociohistorical and Structural Analysis.* Amsterdam: John Benjamins.
Huber, M. 2009. German colonial sources and the history of Pidgin English. A first analysis of the holdings of the Deutsche Kolonialbibliothek. Paper delivered at SPCL Köln, Germany, 11–15 August.
Huber, M. 2013. Ghanaian Pidgin English. In Michaelis et al. (eds), Vol. I.
Huddleston, R. 1994. The contrast between interrogatives and questions. *Journal of Linguistics* 30(2): 411–439.
Huddleston, R. & Pullum, G. 2002. *The Cambridge Grammar of the English Language.* Cambridge: CUP.
Hutchisson, D. 1986. Sursurunga pronouns and the special uses of quadral number. In Wiesemann (ed.), 1–20.
Hyman, L. 2001. Fieldwork as a state of mind. In Newman & Ratliff (eds), 15–33.
Hyman, L. 2006. Word-prosodic typology. *Phonology* 23: 225–257.
Hyman, L. 2008. Universals in phonology. *The Linguistic Review* 25: 83–137.
Iggesen, O. 2011. Number of cases. In Dryer & Haspelmath (eds), Chapter 49. <http://wals.info/chapter/49> (11 May 2011).
Israel, A. & Sandler, W. 2010. Phonological category resolution. A study of handshapes in younger and older Sign Languages. In *Formational Units in Sign Language,* R. Channon & H. van der Hulst (eds), 177–202. Nijmegen: Ishara Press.
Iwasaki, S. & Ingkaphirom, P. 2009. *A Reference Grammar of Thai.* Cambridge: CUP.
Jacob, J. 1968. *Introduction to Cambodian.* Oxford: OUP.
Jacobowitz, E. & Stokoe, W. 1988. Signs of tense in ASL verbs. *Sign Language Studies* 60(Fall): 331–40.
Jakobson, R. 1929. Remarques sur l'évolution phonologique du russe comparée à celle des autres langues slave. *Travaux du Cercle Linguistique de Prague* 2: 7–116.
Jakobson, R. 1958. Typological studies and their contribution to historical comparative linguistics. In *Proceedings of the Eighth International Congress of Linguistic,* 17–35. Oslo: Oslo University Press.
Janda, R. 2001. Beyond "pathways" and "unidirectionality". On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23: 265–340.
Janzen, T. 2007. The expression of grammatical categories in signed languages. In *Verbal and Signed Languages. Comparing Structures, Constructs and Methodologies,* E. Pizzuto, P. Pietrandrea & R. Simone (eds), 171–198. Berlin: Mouton de Gruyter.
Janzen, T., O'Dea, B. & Schaffer, B. 2001. The construal of events. Passives in American Sign Language. *Sign Language Studies* 1(3): 281–310.
Janzen, T. & Shaffer, B. 2002. Gesture as the substrate in the process of ASL grammaticalization. In *Modality and Structure in Signed and Spoken Languages,* R. P. Meier, K. Cormier & D. Quinto-Pozos (eds), 199–224. Cambridge: CUP.
Jendraschek, G. 2006. Basque in contact with Romance languages. In Aikhenvald & Dixon (eds), (a) 143–62.
Jessen, M. & Pétursson, M. 1998. Voiceless nasal phonemes in Icelandic. *Journal of the International Phonetic Association* 28(1&2): 43–53.
Johanson, L. 2000. Viewpoint operators in European languages. In *Tense and Aspect in the Languages of Europe,* Ö. Dahl (ed.), 27–187. Berlin: Mouton de Gruyter.

Johanson, L. 2001. The aspectually neutral situation type. In *Aktionsart and Aspectotemporality in Non-European Languages,* K. H. Ebert & F. Zúñiga (eds), 7–13. Zürich: ASAS Verlag.
Johanson, L. 2008. Remodeling grammar. Copying, conventionalization, grammaticalization. In *Language Contact and Contact Languages,* P. Siemund & N. Kintana (eds), 61–79. Amsterdam: John Benjamins.
Johnson, H. 2000. A Grammar of San Miguel Chimalapa Zoque (Mexico). PhD dissertation, The University of Texas at Austin.
Johnston, R. 1980. *Nakanai of New Britain. The Grammar of an Oceanic Language.* Canberra: The Australian National University.
Johnston, T. 2006. Sign language morphology. In Brown (ed.). 324–328.
Johnston, T. & Schembri, A. 2009. *Australian Sign Language (AUSLAN). An Introduction to Sign Language Linguistics.* Cambridge: CUP.
Joseph, B. 1988. Pronominal affixes in Modern Greek. The case against clisis. In *Papers from the 24th Regional Meeting, Chicago Linguistic Society,* L. MacLeod, G. Larson & D. Brentari (eds), 203–223. Chicago IL: Chicago Linguistic Society.
Joseph, B. 1992. The Balkan languages. In *International Encyclopedia of Linguistics,* W. Bright (ed.), 153–155. Oxford: OUP.
Joseph, B. D. & Janda, R. D. (eds). 2003. *The Handbook of Historical Linguistics.* Oxford: Blackwell.
Joseph, B. & Philippaki-Warburton, I. 1987. *Modern Greek.* London: Croom Helm.
Kakumasu, J. 1968. Urubú Sign Language. *International Journal of American Linguistics* 34(4): 275–281.
Kärkkäinen, E., Sorjonen, M.-L. & Helasvuo, M.-L. 2007. Discourse structure. In Shopen (ed.), Vol. II: 301–420.
Kastenholz, R. 2003. Auxiliaries, grammaticalization, and word order in Mande. *Journal of African Languages and Linguistics* 24: 31–53.
Keegan, J. 1997. *A Reference Grammar of Mbay.* München: Lincom Europa.
Keen, S. 1983. Yukulta. In *Handbook of Australian languages* 3, R. M. W. Dixon & B. J. Blake (eds), 191–304. Amsterdam: John Benjamins.
Keenan, E. 1976. Towards a universal definition of "subject". In *Subject and Topic,* C. N. Li (ed.), 303–333. New York NY: Academic Press.
Keenan, E. & Dryer, M. 2007. Passive in the world's languages. In Shopen (ed.), Vol. I: 325–361.
Keenan, E. & Polinsky, M. 1998. Malagasy (Austronesian). In Spencer & Zwicky (eds), 563–623.
Kenesei, I., Vago, R. & Fenyvesi, A. 1998. *Hungarian.* London: Routledge.
Key, H. 1967. *Morphology of Cayuvava.* Berlin: Mouton de Gruyter.
Khalilova, Z. 2009. *A Grammar of Khwarshi.* Utrecht: LOT.
Kiessling, R. 1994. *Eine Grammatik des Burunge.* Köln: Rüdiger Köppe.
King, G. 1996. *Modern Welsh. A Comprehensive Grammar.* London: Routledge.
Kitching, A. 1915. *A Handbook of the Ateso Language.* London: Society for Promoting Christian Knowledge.
Klaiman, M. 1991. *Grammatical Voice.* Cambridge: CUP.
Klamer, M. 1998. *A Grammar of Kambera.* Berlin: Mouton de Gruyter.
Klein, T. 2013. Gullah. In Michaelis et al. (eds), Vol. I.
Kohler, K. 1990. Illustrations of the IPA. German. *Journal of the International Phonetic Association* 20(1): 48–50.
König, E. & Siemund, P. 2007. Speech act distinctions in grammar. In Shopen (ed.), Vol. I, 276–324.
König, E. & Siemund, P. 2011. Intensifiers and and reflexive pronouns. In Dryer & Haspelmath (eds), Ch. 47. <http://wals.info/chapter/47> (11 May 2011).

Koptjevskaja-Tamm, M. 2010. Linguistic typology and contact linguistics. In Song (ed.), 568–590.
Koptjevskaja-Tamm, M. & Wälchi, B. 2001. The Circum-Baltic languages. An areal-typological approach. In *The Circum-Baltic Languages. Grammar and Typology,* Ö. Dahl & M. Koptjevskaja-Tamm (eds), 615–750. Amsterdam: John Benjamins.
Kornfilt, J. 1990. Turkish and the Turkic languages. In *The World's Major Languages,* B. Comrie (ed.), 619–644. Oxford: OUP.
Kornfilt, J. 2003. *Turkish.* London: Routledge.
Kortmann, B. 1997. *Adverbial Subordination. A Typology and History of Adverbial Subordinators Based on European Languages.* Berlin: Mouton de Gruyter.
Kossmann, M. 1997. *Grammaire du parler berbère de Figuig. (Maroc oriental).* Paris: Peeters.
Kozinsky, I., Nedjalkov, V. & Polinsky, M. 1988. Antipassive in Chukchee. In *Passive and Voice.* M. Shibatani (ed.), 651–706. Amsterdam: John Benjamins.
Krause, T. B. & Slocum, J. 2007–2010. Tocharian online. <http://www.utexas.edu/cola/centers/lrc/eieol/tokol-0-X.html>
Krauss, M. 1992. The world's languages in crisis. *Language* 68(1): 4–10.
Krauss, M. 2007. Keynote: Mass language extinction and documentation. The race against time. In *The Vanishing Languages of the Pacific Rim,* O. Miyaoka, O. Sakiyama & M. E. Krauss (eds), 3–24. Oxford: OUP.
Kulikov, L. 2010. Voice typology. In Song (ed.), 368–398.
Kutsch Lojenga, C. 1994. *Ngiti. A Central Sudanic Language of Zaire.* Köln: Rüdiger Köppe.
Labov, W. 1972. *Sociolinguistic Patterns.* Philadelphia PA: University of Pennsylvania Press.
Ladefoged, P. 1992. Another view of endangered languages. *Language* 68(4): 809–811.
Ladefoged, P. 2005. *Vowels and Consonants. An Introduction to the Sounds of Languages,* 2nd edn. Oxford: Blackwell.
Ladefoged, P. & Johnson, K. 2010. *A Course in Phonetics,* 6th edn. Boston MA: Wadsworth/Cengage Learning.
Ladefoged, P., Ladefoged, J., Turk, A., Hind, K. & Skilton, S. J. 1998. Phonetic structures of Scottish Gaelic. *Journal of the International Phonetic Association* 28(1&2): 1–41.
Ladefoged, P. & Maddieson, I. 1996. *The Sounds of the World's Languages.* Oxford: Blackwell.
Laidig, W. & Laidig, C. 1990. Larike pronouns. Duals and trials in a Central Moluccan language. *Oceanic Linguistics* 29(2): 87–109.
Lambrecht, K. 1994. *Information Structure and Sentence Form. Topic, Focus, and the Mental Representations of Discourse Referents.* Cambridge: CUP.
Landau, E., Lončarić, M., Horga, D. & Škarić, I. 1995. Illustrations of the IPA. Croatian. *Journal of the International Phonetic Association* 25(2): 83–86.
Lane, H. 1984. *When the Mind Hears. A History of the Deaf.* New York NY: Random House.
Lang, J. 2013. Cape Verdean Creole of Santiago. In Michaelis et al. (eds), Vol. II.
LaPolla, R. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In Aikhenvald & Dixon (eds), 225–254.
LaPolla, R. 2003. Evidentiality in Qiang. In *Studies in Evidentiality,* A. Aikhenvald & R. M. W. Dixon (eds), 63–78. Amsterdam: John Benjamins.
LaPolla, R. & Poa, D. 2006. On describing word order. In *Catching Language. The Standing Challenge of Grammar Writing,* F. K. Ameka, A. Dench & N. Evans (eds), 269–98. Berlin: Mouton de Gruyter.
Larsen, B. 1977. *Orokaiva Grammar.* Ms.
Laufer, A. 1990. Illustrations of the IPA. Hebrew. *Journal of the International Phonetic Association* 20(2): 40–43.

Laver, J. 1975. Communicative functions of phatic communion. In *Organization of Behavior in Face-to-Face Interaction,* A. Kendon, R. M. Harris & M. Ritchie Key (eds), 215–240. The Hague: Mouton de Gruyter.

Lee, W.-S. & Zee, E. 2009. Illustrations of the IPA. Hakka Chinese. *Journal of the International Phonetic Association* 39(1): 107–111.

Leeding, V. 1989. Anindilyakwa Phonology and Morphology. PhD dissertation, University of Sydney.

Leger, R. 1991. Sprachproben aus dem Westtschadischen. Kupto- und Kwamitexte. *Afrikanistische Arbeitspapiere* 28: 5–32.

Lehmann, C. 1982. *Thoughts on Grammaticalization,* 1st edn. Erfurt: Seminar für Sprachwissenschaft der Universität.

Lehmann, C. 1984. *Der Relativsatz. Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik.* Tübingen: Narr.

Lehmann, C. 1999. *Documentation of Endangered Languages. A Priority Task for Linguists.* Erfurt: Seminar für Sprachwissenschaft der Universität.

Lehmann, C. 2000. Interlinear morphemic glossing. In Booij et al. (eds), 1834–1857.

Lehmann, C. 2002. *Thoughts on Grammaticalization,* 2nd, revised edn. Erfurt: Seminar für Sprachwissenschaft der Universität.

Lehmann, C. & Moravcsik, E. 2000. Noun. In Booij et al. (eds), 732–757.

Levinson, S. 1983. *Pragmatics.* Cambridge: CUP.

Lewis, M. P. (ed.). 2009. *Ethnologue. Languages of the world.* Dallas TX: SIL. <http://www.ethnologue.com>

Li, C. & Thompson, S. 1990. Chinese. In *The World's Major Languages,* B. Comrie (ed.), 811–833. Oxford: OUP.

Lichtenberk, F. 1985. Syntactic-category change in Oceanic languages. *Oceanic Linguistics* 24(1-2): 1–84.

Lichtenberk, F. 1991. On the gradualness of grammaticalization. In *Approaches to Grammaticalization.* Vol. I: *Focus on Theoretical and Methodological Issues,* E. C. Traugott & B. Heine (eds), 37–80. Amsterdam: John Benjamins.

Lichtenberk, F. 2005. On the notion 'adjective' in Toqabaqita. *Oceanic Linguistics* 44(1): 113–144.

Lichtenberk, F. 2006. Serial verb constructions in Toqabaqita. In *Serial Verb Constructions. A Cross-Linguistic Typology,* A. Aikhenvald & R. M. W. Dixon (eds), 254–272. Oxford: OUP.

Lichtenberk, F. 2007. A typologically unusual interrogative word in Toqabaqita and other Oceanic languages. *Oceanic Linguistics* 46(2): 603–612.

Liddell, S. 1984. THINK and BELIEVE. Sequentiality in American Sign Language. *Language* 60(2): 372–399.

Liddell, S. 2003. *Grammar, Gesture, and Meaning in American Sign Language.* Cambridge: CUP.

Longacre, R. 2007. Sentence as combinations of clauses. In Shopen (ed.), Vol. II: 372–420.

Lord, C. 1993. *Historical Change in Serial Verb Constructions.* Amsterdam: John Benjamins.

Lucy, J. 1992. *Grammatical Categories and Cognition. A Case Study of the Linguistic Relativity Hypothesis.* Cambridge: CUP.

Lynch, J. 2002. Cèmuhî. In Lynch et al. (eds), 753–64.

Lynch, J., Ross, M. & Crowley, T. (eds). 2002. *The Oceanic Languages.* Richmond: Curzon.

Macaulay, M. 1996. *A Grammar of Chalcatongo Mixtec.* Berkeley CA: University of California Press.

MacDonald, L. 1990. *A Grammar of Tauya.* Berlin: Mouton de Gruyter.

Maddieson, I. 1984. *Patterns of Sounds.* Cambridge: CUP.

Maddieson, I. 1989. Aerodynamic constraints on sound change. The case of bilabial trills. In *Working Papers in Phonetics,* No. 72, 91–115. Los Angeles CA: Department of Linguistics, UCLA.

Maddieson, I. 2010. Typology of phonological systems. In Song (ed.), 534–48.
Maddieson, I. 2011a. Absence of common consonants. In Dryer & Haspelmath (eds), Ch. 18. <http://wals.info/chapter/18> (11 May 2011).
Maddieson, I. 2011b. Consonant inventories. In Dryer & Haspelmath (eds), Ch.1, <http://wals.info/chapter/1> (11 May 2011).
Maddieson, I. 2011c. Front rounded vowels. In Dryer & Haspelmath (eds), Ch.11. <http://wals.info/chapter/11> (11 May 2011).
Maddieson, I. 2011d. Presence of uncommon consonants. In Dryer & Haspelmath (eds), Ch. 19. <http://wals.info/chapter/19> (11 May 2011).
Maddieson, I. 2011e. Syllable structure. In Dryer & Haspelmath (eds). chapter 12. Available from: http://wals.info/chapter/12 (11 May 2011).
Maddieson, I. 2011f. Tone. In Dryer & Haspelmath (eds), Ch. 13. <http://wals.info/chapter/13> (11 May 2011).
Maddieson, I. 2011g. Uvular consonants. In Dryer & Haspelmath (eds), Ch. 6. <http://wals.info/chapter/6> (11 May 2011).
Maddieson, I. 2011h. Vowel quality inventories. In Dryer & Haspelmath (eds), Ch. 2. <http://wals.info/chapter/2> (11 May 2011).
Mahootian, S. 1997. *Persian.* London: Routledge.
Malchukov, A., Haspelmath, M. & Comrie, B. 2010a. Ditransitive constructions. A typological overview. In Malchukov et al. (eds), 1–64.
Malchukov, A., Haspelmath, M. & Comrie, B. (eds). 2010b. *Studies in Ditransitive Constructions. A Comparative Handbook.* Berlin: Mouton de Gruyter.
Manfredi, S. & Petrollino, S. 2013. Juba Arabic. In Michaelis et al. (eds), Vol. III.
Marantz, A. 1982. Re reduplication. *Linguistic Inquiry* 13(3): 435–482.
Marchese, L. 1976. Subordination in Godié. MA thesis, University of California, Los Angeles.
Marchese, L. 1986. *Tense, Aspect and the Development of Auxiliaries in Kru Languages.* Dallas TX: SIL.
Marlett, S. 2008. Stress, extrametricality and the minimal word in Seri. *Linguistic Discovery* 6(1): 1–14.
Marnita, R. 1996. Classifiers in Minangkabau. MA thesis, Australian National University.
Marsh, J. 1969. Mantjiltjara phonology. *Oceanic Linguistics* 8(2): 131–152.
Martins, S. 1994. Análise da morfossintaxe da lingua Dâw (Maku-Kamã) e sua classificação. MA thesis, Universidade Federal de Santa Catarina.
Martins, S. & Martins, V. 1999. Makú. In *The Amazonian Languages,* R. M. W. Dixon & A. Aikhenvald (eds), 251–268. Cambridge: CUP.
Maslova, E. 2003. *A Grammar of Kolyma Yukaghir.* Berlin: Mouton de Gruyter.
Maslova, E. & Nedjalkov, V. 2011. Reciprocal constructions. In Dryer & Haspelmath (eds), Ch. 106. <http://wals.info/chapter/106> (11 May 2011).
Massam, D. 2009. Noun incorporation. Essentials and extensions. *Language and Linguistics Compass* 3(4): 1075–1096.
Massone, M. & Curiel, M. 2004. Sign order in Argentine Sign Language. *Sign Language Studies* 5(1): 63–93.
Matras, Y. 2007. The borrowability of structural features. In Matras & Sakel (eds), 31–73.
Matras, Y. & Bakker, P. 2003. The study of mixed languages. In *The Mixed Language Debate. Theoretical and Empirical Advances,* Y. Matras & P. Bakker (eds), 1–21. Berlin: Mouton de Gruyter.
Matras, Y. & Sakel, J. (eds). 2007. *Grammatical Borrowing in Cross-Linguistic Perspective.* Berlin: Mouton de Gruyter.

Matthews, P. 2006. *Morphology,* 2nd edn. Cambridge: CUP.
Matthews, S. & Li, M. 2013. Chinese Pidgin English. In Michaelis et al. (eds), Vol. I.
Matthews, S. & Yip, V. 1994. *Cantonese. A Comprehensive Grammar.* London: Routledge.
Maurer, P. 2013a. Angolar. In Michaelis et al. (eds), Vol. II.
Maurer, P. 2013b. Batavia Creole. In Michaelis et al. (eds), Vol. II.
Maurer, P. 2013c. Principense. In Michaelis et al. (eds), Vol. II.
McCarthy, J. & Broselow, E. 1983. A theory of internal reduplication. *Linguistic Review* 3(1): 25–88.
McGregor, W. 2009. Typology of ergativity. *Language and Linguistics Compass* 3(1): 480–508.
McGuckin, C. 2002. Gapapaiwa. In Lynch et al. (eds), 297–321.
Mchombo, S. 1998. Chichewa (Bantu). In Spencer & Zwicky (eds), 500–520.
McKaughan, H. 1958. *The Inflection of Syntax of Maranao Verbs.* Manila: Institute for National Language.
McKee, D., McKee, R. & Major, G. 2008. Sociolinguistic variation in NZSL numerals. In Müller de Quadros (ed.), 296–313.
McLendon, S. 1978. Ergativity, case and transitivity in Eastern Pomo. *International Journal of American Linguistics* 44(1): 1–9.
Meakins, F. 2013. Gurindji Kriol. In Michaelis et al. (eds), Vol. III.
Meeuwis, M. 2013. Lingala. In Michaelis et al. (eds), Vol. III.
Meinhof, C. 1930. *Der Koranadialekt des Hottentottischen.* Berlin: Dietrich Reimer.
Meir, I. 1999. A perfect marker in Israeli Sign Language. *Sign Language & Linguistics* 2(1): 41–60.
Meir, I. 2002. A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory* 20: 413–450.
Meir, I. 2004. Question and negation in Israeli Sign Language. *Sign Language & Linguistics* 7(2): 97–124.
Meir, I. 2010. The emergence of argument structure in two new Sign Languages. In *Syntax, Lexical Semantics and Event Structure,* M. Rappaport Hovav, E. Doron & I. Sichel (eds), 101–123. Oxford: OUP.
Meira, S. 1999. A Grammar of Tiriyo (Brazil, Suriname). PhD dissertation, Rice University.
Meira, S. & Gildea, S. 2009. Property concepts in the Cariban family. Adjectives, adverbs, and/or nouns?. In *The Linguistics of Endangered Languages. Contributions to Morphology and Morphosyntax,* W. L. Wetzels (ed.), 95–134. Utrecht: LOT.
Merlan, F. 1976. Noun incorporation and discourse reference in Modern Nahuatl. *International Journal of American Linguistics* 42(3): 177–191.
Merlan, F. 1994. *A Grammar of Wardaman. A Language of the Northern Territory of Australia.* Berlin: Mouton de Gruyter.
Merrill, E. 2008. Illustrations of the IPA. Tilquiapan Zapotec. *Journal of the International Phonetic Association* 38(1): 107–114.
Mesthrie, R & Surek-Clark, C. 2013. Fanakalo. In Michaelis et al. (eds), Vol. III.
Meyerhoff, M. 2006. *Introducing Sociolinguistics.* London: Routledge.
Meyerhoff, M. 2013. Bislama. In Michaelis et al. (eds), Vol. I.
Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013a. *Survey of Pidgin and Creole Languages,* Vol. I: *English-based and Dutch-based Languages.* Oxford: OUP.
Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013b. *Survey of Pidgin and Creole Languages,* Vol. II: *Portuguese-based, Spanish-based and French-based Languages.* Oxford: OUP.
Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds). 2013c. *Survey of Pidgin and Creole Languages,* Vol. III: *Pidgins, Creoles and Mixed Languages Based on Languages from Africa, Asia, Australia and the Americas.* Oxford: OUP.

Michaelis, S., Maurer, P., Haspelmath, M. & Huber, M. (eds) and the APiCS Consortium (contr.). 2013. *Atlas of Pidgin and Creole Language Structures Online.* München: MPDL.
Miestamo, M. 2007. Negation. An overview of typological research. *Language and Linguistics Compass* 1(5): 552–570.
Miestamo, M. 2011a. Subtypes of asymmetric standard negation. In Dryer & Haspelmath (eds), Ch. 114. <http://wals.info/chapter/114> (11 May 2011).
Miestamo, M. 2011b. Symmetric and asymmetric standard negation. In Dryer & Haspelmath (eds), Ch. 113. <http://wals.info/chapter/113> (11 May 2011).
Migge, B. 2013. Nengee. In Michaelis et al. (eds), Vol. I.
Miller, A. 2001. *A Grammar of Jamul Tiipay.* Berlin: Mouton de Gruyter.
Milne, A. 1974. *Winnie-the-Pooh.* London: Methuen Children's Books.
Miner, K. 1979. Dorsey's law in Winnebago-Chiwere and Winnebago accent. *International Journal of American Linguistics,* 45(1): 25–33.
Mithun, M. 1984. The evolution of noun incorporation. *Language* 60(4): 847–894.
Mithun, M. 1988. The grammaticalization of coordination. In *Clause Combining in Grammar and Discourse,* J. Haiman & S. A. Thompson (eds), 331–359. Amsterdam: John Benjamins.
Mithun, M. 1992. Is basic word order universal? In *Pragmatics of Word Order Flexibility,* D. L. Payne (ed.), 15–62. Amsterdam: John Benjamins.
Mithun, M. 1993. 'Switch-reference'. Clause combining in Central Pomo. *International Journal of American Linguistics* 59(2): 119–136.
Miyaoka, O., Sakiyama, O. & Krauss, M. E. (eds). 2007. *The Vanishing Languages of the Pacific Rim.* Oxford: OUP.
Moore, B. & Franklin, G. 1979. *Breves notícias da língua Maku-Hupda.* Brasília: SIL.
Moravcsik, E. 2003. A semantic analysis of associative plurals. *Studies in Language* 27(3): 469–503.
Moravcsik, E. 2010. Explaining language universals. In Song (ed.), 69–89.
Mufwene, S. 2013. Kikongo-Kituba. In Michaelis et al. (eds), Vol. III.
Muller, E. 1989. Inleiding tot de sintaxis van het Papiamentu. PhD dissertation, University of Amsterdam.
Müller de Quadros, R. 2008. Back to back(wards) and moving on. On agreement, auxiliaries and verb classes in Sign Languages. In Müller de Quadros (ed.), 530–551.
Müller de Quadros, R. (ed.). 2008. *Sign Languages. Spinning and Unraveling the Past, Present and Future* [TISLR 9]. Petrópolis: Editoria Arara Azul.
Muravyova, I. 1998. Chukchee (Paleo-Siberian). In Spencer & Zwicky (eds), 521–538.
Muysken, P. 2013. Media Lengua. In Michaelis et al. (eds), Vol. III.
Muysken, P. & Smith, N. 1995. The study of pidgin and creole languages. In *Pidgins and Creoles. An Introduction,* J. Arends, P. C. Muysken & N. Smith (eds), 3–14. Amsterdam: John Benjamins.
Myers-Scotton, C. 2002. *Contact Linguistics. Bilingual Encounters and Grammatical Outcomes.* Oxford: OUP.
Myhill, J. 1994. A quantitative analysis of the use of the Javanese speech levels in a modern anthology. *Oceanic Linguistics* 33(1): 73–103.
Narang, G. & Becker, D. 1971. Aspiration and nasalization in the generative phonology of Hindi-Urdu. *Language* 47(3): 646–667.
Narrog, H. & Heine, B. 2011. *The Oxford Handbook of Grammaticalization*. Oxford: OUP.
Nathan, G. 2001. Clicks in a Chinese nursery rhyme. *Journal of the International Phonetic Association* 31(2): 223–8.
Nedjalkov, I. 1997. *Evenki.* London: Routledge.

Nespor, M. & Sandler, W. 1999. Prosody in Israeli Sign Language. *Language and Speech* 42(2-3): 143–176.
Nettle, D. & Romaine, S. 2002. *Vanishing Voices. The Extinction of the World's Languages.* Oxford: OUP.
Newman, P. & Ratliff, M. S. (eds). 2001. *Linguistic Fieldwork.* Cambridge: CUP.
Ní Chasaide, A. 1999. Illustrations of the IPA. Irish. In *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet,* 111–117. Cambridge: CUP.
Nichols, J. 1986. Head-marking and dependent-marking grammar. *Language* 62(1): 56–119.
Nichols, J. 2007. What, if anything, is typology? *Linguistic Typology* 11: 231–238.
Nichols, J. & Bickel, B. 2011a. Locus of marking in possessive noun phrases. In Dryer & Haspelmath (eds), Ch. 24. <http://wals.info/chapter/24> (11 May 2011).
Nichols, J. & Bickel, B. 2011b. Obligatory possessive inflection. In Dryer & Haspelmath (eds), Ch. 58. <http://wals.info/chapter/58> (11 May 2011).
Nielsen, N. 1989. *Dansk etymologisk ordbog. Ordenes historie,* 3rd edn. Copenhagen: Gyldendal.
Nikitina, T. 2008. Locative terms and spatial frames of reference in Wan. *Journal of African Languages and Linguistics* 29: 29–47.
Nikolaeva, I. 2001. Secondary topic as a relation in information structure. *Linguistics* 39(1): 1–49.
Nikolaeva, I. & Tolskaya, M. 2001. *A Grammar of Udihe.* Berlin: Mouton de Gruyter.
Nonaka, A. 2004. The forgotten endangered languages. Lessons on the importance of remembering from Thailand's Ban Khor Sign Language. *Language in Society* 33(5): 737–767.
Noonan, M. 1992. *A Grammar of Lango.* Berlin: Mouton de Gruyter.
Noonan, M. 2003. Chantyal. In *The Sino-Tibetan Languages,* G. Thurgood & R. J. LaPolla (eds), 315–335. London: Routledge.
Noonan, M. 2007. Complementation. In Shopen (ed.), Vol. II: 52–150.
Norde, M. 2001. Deflexion as a counterdirectional factor in grammatical change. *Language Sciences* 23: 231–264.
Nordlinger, R. & Sadler, L. 2004. Nominal tense in crosslinguistic perspective. *Language* 80(4): 776–806.
Nurse, D. 2008. *Tense and Aspect in Bantu.* Oxford: OUP.
Nyst, V. 2007. *A Descriptive Analysis of Adamorobe Sign Language (Ghana).* Utrecht: LOT.
Ó Dochartaigh, C. 1992. The Irish language. In *The Celtic Languages,* D. MacAulay (ed.), 11–99. Cambridge: CUP.
Oda, S. 1977. The Syntax of Pulo Annian. A Nuclear Micronesian Language. PhD dissertation, University of Hawai'i at Manoa.
Odden, D. 1996. *The Phonology and Morphology of Kimatuumbi.* Oxford: OUP.
Ohala, M. 1994. Illustrations of the IPA. Hindi. *Journal of the International Phonetic Association* 24(1): 35–38.
Olawsky, K. 2006. *A Grammar of Urarina.* Berlin: Mouton de Gruyter.
Olson, M. 1981. Barai Clause Junctures. Toward a functional theory of interclausal relations. PhD dissertation, Australian National University.
Osumi, M. 1995. *Tinrin Grammar.* Honolulu HI: University of Hawai'i Press.
Owens, J. 1985. *A Grammar of Harar Oromo (Northeastern Ethiopia). Including a Text and a Glossary.* Hamburg: Helmut Buske.
Paauw, S. 2013. Ambon Malay. In Michaelis et al. (eds), Vol. III.
Palmer, F. 2001. *Mood and Modality,* 2nd edn. Cambridge: CUP.
Pans, A. 1998. The mother-in-law taboo. *Ethnology* 37(1): 71–97.
Parker, S. 1999. On the behavior of definite articles in Chamicuro. *Language* 75(3): 552–562.

Payne, D. 1985a. Aspects of the Grammar of Yagua: A Typological Perspective (Peru). PhD dissertation, University of California, Los Angeles.
Payne, J. 1980. The decay of ergativity in Pamir languages. *Lingua* 51: 147–86.
Payne, J. 1985b. Complex phrases and complex sentences. In *Language Typology and Syntactic Description*. Vol II: *Complex Constructions*, T. Shopen (ed.), 3–41. Cambridge: CUP
Payne, J. 1989. Pāmir languages. In *Compendium linguarum Iranicarum,* R. Schmitt (ed.), 417–44. Wiesbaden: Reichert.
Payne, T. 1997. *Describing Morphosyntax. A Guide for Field Linguists.* Cambridge: CUP.
Payne, T. E. & Weber, D. (eds). 2007. *Perspectives on Grammar Writing.* Amsterdam: John Benjamins.
Pearsall, J. & Trumble, B. 1996. *The Oxford English Reference Dictionary,* 2nd edn. Oxford: OUP.
Perekhvalskaya, E. 2013. Chinese Pidgin Russian. In Michaelis et al. (eds), Vol. III.
Pericliev, V. 2008. Implicational phonological universals. *Folia Linguistica* 42(1): 195–225.
Perlmutter, D. 1992. Sonority and syllable structure in American Sign Language. *Linguistic Inquiry* 23(3): 407–442.
Perniss, P., Pfau, R. & Steinbach, M. 2007. Can't you see the difference? Sources of variation in sign language structure. In *Visible Variation: Cross-Linguistic Studies on Sign Language Structure,* P. Perniss, R. Pfau & M. Steinbach (eds), 1–34. Berlin: Mouton de Gruyter.
Peterson, D. 2001. Ingush ʔa. The elusive type 5 clitic?. *Language* 77(1): 144–155.
Peterson, D. 2007. *Applicative Constructions.* Oxford: OUP.
Pfau, R. & Quer, J. 2010. Nonmanuals. Their prosodic and grammatical roles. In *Sign Languages,* D. Brentari (ed.), 381–402. Cambridge: CUP.
Pfau, R. & Steinbach, M. 2006a. *Modality-independent and Modality-specific Aspects of Grammaticalization in Sign Languages.* Potsdam: Universitäts Verlag Potsdam.
Pfau, R. & Steinbach, M. 2006b. Pluralization in sign and in speech. A cross-modal typological study. *Linguistic Typology* 10: 135–82.
Pitkin, H. 1985. *Wintu Grammar.* Berkeley CA: University of California Press.
Pitman, D. 1980. *Bosquejo de la Gramatica araona* [Notas Linguisticas No. 9]. Riberalto, Bolivia: Instituto Lingüístico de Verano.
Pizzuto, E. & Corazza, S. 1996. Noun morphology in Italian Sign Language (LIS). *Lingua* 98: 169–196.
Poedjosoedarmo, S. 1968. Javanese speech levels. *Indonesia* 6: 54–81.
Polinsky, M. 2011a. Antipassive constructions. In Dryer & Haspelmath (eds), Ch. 108. <http://wals.info/chapter/108> (11 May 2011).
Polinsky, M. 2011b. Applicative constructions. In Dryer & Haspelmath (eds), Ch. 109. <http://wals.info/chapter/109> (11 May 2011).
Poppe, N. 1964. *Bashkir Manual.* Bloomington IN: Indiana University Press.
Post, M. 2013. Fa d'Ambô. In Michaelis et al. (eds), Vol. II.
Press, M. 1975. A Grammar of Chemehuevi. PhD dissertation, University of California, Los Angeles.
Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language.* Harlow: Longman.
Quizar, R. & Knowles-Berry, S. 1988. Ergativity in the Cholan languages. *International Journal of American Linguistics* 54(1): 73–95.
Ramat, P. 2010. The (early) history of linguistic typology. In Song (ed.), 9–24.
Rankin, R. 1982. A Quapaw vocabulary. *Kansas Working Papers in Linguistics* 7: 125–152.
Rapold, C. 2006. Towards a Grammar of Benchnon. PhD disseration, Universiteit Leiden.
Regueira, X. 1996. Illustrations of the IPA. Galician. *Journal of the International Phonetic Association* 26(2): 119–122.
Reh, M. 1985. *Die Krongo-Sprache (Niinò Mó-Dì).* Berlin: Dietrich Reimer.

Reid, L. 1992. On the development of the aspect system in some Philippine languages. *Oceanic Linguistics* 31(1): 65–91.
Remijsen, B. & Adong Manyang, C. 2009. Illustrations of the IPA. Luanyjang Dinka. *Journal of the International Phonetic Association* 39(1): 113–124.
Rennison, J. 1997. *Koromfe.* London: Routledge.
Ridouane, R. 2008. Syllables without vowels. Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology* 25: 321–359.
Rijkhoff, J. 2007. Word classes. *Language and Linguistics Compass* 1(6): 709–726.
Rijkhoff, J. & Bakker, D. 1998. Language sampling. *Linguistic Typology* 2: 263–314.
Roberts, J. 1988. Amele switch-reference and the theory of grammar. *Linguistic Inquiry* 19(1): 45–63.
Roberts, J. 1990. Modality in Amele and other Papuan languages. *Journal of Linguistics* 26: 363–401.
Roberts, S. 2000. Nativization and the genesis of Hawaiian Creole. In *Language Change and Language Contact in Pidgins and Creoles,* J. H. McWhorter (ed.), 257–300. Amsterdam: John Benjamins.
Robinson, S. 2006. The phoneme inventory of the Aita dialect of Rotokas. *Oceanic Linguistics* 45(1): 206–209.
Romaine, S. 2007. Preserving endangered languages. *Language and Linguistics Compass* 1(1-2): 115–32.
Romaine, S. & Lange, D. 1991. The use of *like* as a marker of reported speech and thought. A case of grammaticalization in progress. *American Speech* 66(3): 227–279.
Romero-Figueroa, A. 1985. OSV as the basic order in Warao. *Linguistics* 23: 105–121.
Rood, D. 1996. Sketch of Wichita, a Caddoan language. In *Handbook of North American Indians,* Vol. 17: *Languages,* I. Goddard (ed.), 580–608. Washington DC: Smithsonian Institute.
Ross, M. 2002a. Bali-Vitu. In Lynch et al. (eds), 362–386.
Ross, M. 2002b. Takia. In Lynch et al. (eds), 216–249.
Routamaa, J. 1994. Kamula Grammar Essentials. Ms. <http://www.sil.org/pacific/png/pubs/50209/Kamula_gram_essent.pdf>
Rubino, C. 2000. *Ilocano Dictionary and Grammar. Ilocano-English, English-Ilocano.* Honolulu HI: University of Hawai'i Press.
Rubino, C. 2011. Reduplication. In Dryer & Haspelmath (eds), Ch. 27. <http://wals.info/chapter/27> (11 May 2011).
Rude, N. 1985. Studies in Nez Perce Grammar and Discourse (Promotion, Genitives, Antipassive, Objects, Ergative, Indians, Northwest). PhD dissertation, University of Oregon.
Ruge, H. 1984. *Nygrekisk språklära.* Lund: Studentlitteratur.
Rule, W. 1977. *A Comparative Study of the Foe, Huli and Pole Languages of Papua New Guinea.* Sydney: University of Sydney.
Rumsey, A. 1978. A Grammar of Ungarinjin With Special Reference To the Structure of Discourse. PhD dissertation, The University of Chicago.
Ryding, K. 2005. *A Reference Grammar of Modern Standard Arabic.* Cambridge: CUP.
Sadock, J. 2006. Speech acts. In *The Handbook of Pragmatics,* L. R. Horn & G. Ward (eds), 53–73. Malden MA: Blackwell.
Sakel, J. 2007. Types of loan. Matter and pattern. In Matras & Sakel (eds), 15–29.
Sakel, J. & Everett, D. 2012. *Linguistic Fieldwork. A Student Guide.* Cambridge: CUP.
Sakel, J. & Matras, Y. 2008. Modelling contact-induced change in grammar. In *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Rrocesses,* T. Stolz, D. Bakker & R. Salas Palomo (eds), 63–89. Berlin: Mouton de Gruyter.

Samarin, W. 2013. Sango. In Michaelis et al. (eds), Vol. III.
Sampson, G., Gil, D. & Trudgill, P. (eds). 2009. *Language Complexity as an Evolving Variable.* Oxford: OUP.
Sandler, W. 1993. A sonority cycle in American Sign Language. *Phonology* 10: 243–279.
Sandler, W. 2003. Sign language phonology. In *The Oxford International Encyclopedia of Linguistics,* 2nd edn, W. J. Frawley (ed.). Oxford: OUP.
Sandler, W. 2006. Sign Language. Overview. In Brown (ed.), 328–38.
Sandler, W. & Lillo-Martin, D. 2006. *Sign Language and Linguistic Universals.* Cambridge: CUP.
Sapir, E. 1921. *Language. An Introduction to the Study of Speech.* New York NY: Harcourt Brace.
Schachter, P. & Otanes, F. 1983. *Tagalog Reference Grammar.* Berkeley CA: University of California Press.
Schachter, P. & Shopen, T. 2007. Parts-of-speech systems. In Shopen (ed.), Vol. I: 1–60.
Schalber, K. & Grose, D. 2008. The semantics, syntax and phonology of adverbial nonmanuals in Austrian and American Sign Language. In Müller de Quadros (ed.), 552–65.
Schaub, W. 1985. *Babungo.* London: Croom Helm.
Schladt, M. 2000. The typology and grammaticalization of reflexives. In *Reflexives. Forms and Functions,* Z. Frajzyngier & T. S. Curl (eds), 103–124. Amsterdam: John Benjamins.
Schlichter, A. 1986. The origins and deictic nature of Wintu evidentials. In *Evidentiality. The Linguistic Coding of Epistemology,* W. L. Chafe & J. Nichols (eds), 46–59. Norwood NJ: Ablex.
Schultze-Berndt, E. 2000. Simple and Complex Verbs in Jaminjung. A Study of Event Categorization in an Australian Language. PhD dissertation, Katholieke Universiteit Nijmegen.
Schultze-Berndt, E. 2007. Recent grammatical borrowing into an Australian Aboriginal language. The case of Jaminjung and Kriol. In Y. Matras & J. Sakel (eds). 363–86.
Schwager, W. & Zeshan, U. 2008. Word classes in Sign Languages. *Studies in Language* 32(3): 509–545.
Schwartz, J.-L., Boe, L.-J., Vallée, N. & Abry, C. 1997. Major trends in vowel system inventories. *Journal of Phonetics* 25: 233–53.
Schwartz, L. 2000. Pronoun and article. In Booij et al. (eds), 783–94.
Schwartz, L. & Dunnigan, T. 1986. Pronouns and pronominal categories in Southwestern Ojibwe. In Wiesemann (ed.), 285–322.
Seyoum, M. 2008. *A Grammar of Dime.* Utrecht: LOT.
Seyoum, M. 2009. Negation in Dime. In *The Linguistics of Endangered Languages. Contributions to Morphology and Morphosyntax,* W. L. Wetzels (ed.), 189–198. Utrecht: LOT.
Shaw, P. 1985. Modularisation and substantive constraints in Dakota lexical phonology. *Phonology Yearbook* 2: 173–202.
Shibatani, M. 1990. *The Languages of Japan.* Cambridge: CUP.
Shibatani, M. & Bynon, T. (eds). 1995. *Approaches to Language Typology.* Oxford: OUP.
Shopen, T. (ed.). 2007a. *Language Typology and Syntactic Description,* Vol. I: *Clause Structure,* 2nd edn. Cambridge: CUP.
Shopen, T. (ed.). 2007b. *Language Typology and Syntactic Description,* Vol. II: *Grammatical Categories and the Lexicon,* 2nd edn. Cambridge: CUP.
Shopen, T (ed.). 2007c. *Language Typology and Syntactic Description,* Vol. III: *Grammatical Categories and the Lexicon,* 2nd edn. Cambridge: CUP.
Siegel, J. 2013. Pidgin Hindustani. In Michaelis et al. (eds), Vol. III.
Siewierska, A. 1984. *The Passive. A Comparative Linguistic Analysis.* London: Croom Helm.
Siewierska, A. 2011a. Alignment of verbal person marking. In Dryer & Haspelmath (eds), Ch. 100. <http://wals.info/chapter/100> (11 May 2011).

Siewierska, A. 2011b. Gender distinctions in independent personal pronouns. In Dryer & Haspelmath (eds), Ch. 44. <http://wals.info/chapter/44> (11 May 2011).
Siewierska, A. 2011c. Passive constructions. In Dryer & Haspelmath (eds), Ch. 107. <http://wals.info/chapter/107> (11 May 2011).
Siewierska, A. 2011d. Verbal person marking. In Dryer & Haspelmath (eds), Ch. 102. <http://wals.info/chapter/102> (11 May 2011).
Silverstein, M. 1972. Chinook Jargon. Language contact and the problem of multi-level generative systems, I. *Language* 48(2): 378–406.
Sippola, E. 2013a. Cavite Chabacano. In Michaelis et al. (eds), Vol. II.
Sippola, E. 2013b. Ternate Chabacano. In Michaelis et al. (eds), Vol. II.
Skalmowski, W. 1974. Transitive verb constructions in the Pamir and Dardic languages. *Studia Indoeuropejskie, Polska Akademia Nauk* 37: 205–212.
Slomanson, P. 2013. Sri Lankan Malay. In Michaelis et al. (eds), Vol. III.
Smith, C. 1997. *The Parameter of Aspect,* 2nd edn. Dordrecht: Kluwer.
Smith, G. & Siegel, J. 2013. Tok Pisin. In Michaelis et al. (eds), Vol. I.
Sneddon, J. 1996. *Indonesian. A Comprehensive Grammar.* London: Routledge.
Soares, M. 1992. O suprassegmental em tikuna e a teoria fonológica. PhD dissertation, Universidade Estadual de Campinas.
Soe, M. 1999. A Grammar of Burmese. PhD dissertation, University of Oregon.
Sohn, H.-M. 2001. *The Korean Language.* Cambridge: CUP.
Solnit, D. 1997. *Eastern Kayah Li. Grammar, Texts, Glossary.* Honolulu HI: University of Hawai'i Press.
Sommer, B. 1970. An Australian language without CV syllables. *International Journal of American Linguistics* 36(1): 57–8.
Song, J. 1996. *Causatives and Causation. A Universal Typological Perspective.* London: Longman.
Song, J. 2001. *Linguistic Typology. Morphology and Syntax.* Harlow: Longman.
Song, J. 2009. Word order patterns and principles. An overview. *Language and Linguistics Compass* 3(5): 1328–1341.
Song, J. 2010. Word order typology. In Song (ed.), 253–279.
Song, J. 2011a. Nonperiphrastic causative constructions. In Dryer & Haspelmath (eds), Ch. 111. <http://wals.info/chapter/111> (11 May 2011).
Song, J. 2011b. Periphrastic causative constructions. In Dryer & Haspelmath (eds), Ch. 110. <http://wals.info/chapter/110> (11 May 2011).
Song, J. (ed.). 2010. *The Oxford Handbook of Linguistic Typology.* Oxford: OUP.
Sonnenschein, A. 2004. A Descriptive Grammar of San Bartolome Zoogocho Zapotec (Mexico). PhD dissertation, University of Southern California.
Spencer, A. 1996. *Phonology. Theory and Description.* Oxford: Blackwell.
Spencer, A. 2008. Does Hungarian have a case system? In *Case and Grammatical Relations. Studies in Honor of Bernard Comrie,* G. G. Corbett & M. Noonan (eds), 35–57. Amsterdam: John Benjamins.
Spencer, A. & Zwicky, A. M. (eds). 1998. *The Handbook of Morphology.* Oxford: Blackwell.
Sridhar, S. 1990. *Kannada.* London: Routledge.
Stasch, R. 2003. Separateness as relation. The iconicity, univocality and creativity of Korowai mother-in-law avoidance. *Journal of the Royal Anthropological Institute* 9(2): 317–337.
Stassen, L. 2000. AND-languages and WITH-languages. *Linguistic Typology* 4: 1–54.
Stassen, L. 2011. Noun phrase conjunction. In Dryer & Haspelmath (eds), Ch. 63. <http://wals.info/chapter/63> (11 May 2011).

Steever, S. 1998. *The Dravidian Languages.* London: Routledge.
Steinkrüger, P. 2013. Zamboanga Chabacano. In Michaelis et al. (eds), Vol. II.
Stirling, L. 1993. *Switch-Reference and Discourse Representation.* Cambridge: CUP.
Stokoe, W. 1960. *Sign Language Structure. An Outline of the visual Communication Systems of the American Deaf.* Silver Spring MD: Linstok Press.
Storch, A. 2006. How long do linguistic areas last? Western Nilotic grammars in contact. In Aikhenvald & Dixon (eds), (a) 94–113.
Stump, G. 1998. Inflection. In Spencer & Zwicky (eds), 13–43.
Suárez, J. 1983. *The Mesoamerican Indian Languages.* Cambridge: CUP.
Suchada, G. 2004. Plang Grammar as Spoken in Huay Namkhun Village, Chang Rai Province. MA thesis, Mahidol University.
Sulkala, H. & Karjalainen, M. 1992. *Finnish.* London: Routledge.
Sunee, K. 2003. Syntactic characteristics of Kasong. An endangered language of Thailand. *Mon-Khmer Studies* 33: 167–182.
Supalla, T. 1990. Serial verbs of motion in ASL. In *Theoretical Issues in Sign Language Research,* S. D. Fischer & P. A. Siple (eds), 127–52. Chicago IL: University of Chicago Press.
Šuštaršič, R., Komar, S. & Petek, B. 1995. Illustrations of the IPA. Slovene. *Journal of the International Phonetic Association* 25(2): 86–90.
Sutton-Spence, R. & Woll, B. 1999. *The Linguistics of British Sign Language. An Introduction.* Cambridge: CUP.
Tadmor, U. 2007. Grammatical borrowing in Indonesian. In Matras & Sakel (eds), 301–328.
Tallerman, M. 2005. *Understanding Syntax,* 2nd edn. London: Hodder Education.
Taylor, C. 1985. *Nkore-Kiga.* Dover NH: Croom Helm.
Tegey, H. 1978. *The Grammar of Clitics. Evidence from Pashto (Afghani) and Other Languages.* Kabul: International Center for Pashto Studies.
Tennant, R. & Brown, M. 1998. *The American Sign Language Handshape Dictionary.* Washington DC: Clerc Books.
Terrill, A. 1999. A Grammar of Lavukaleve. A Papuan Language of the Solomon Islands. PhD dissertation, Australian National University.
Tervoort, B. 1953. *Structurele analyse van visueel taalgebruik binnen en groep dove kinderen*, 2 Vols. Amsterdam: North-Holland.
Thelwall, R. & Sa'Adeddin, M. 1990. Illustrations of the IPA. Arabic. *Journal of the International Phonetic Association* 20(2): 37–39.
Thomas, E. 1978. *A Grammatical Description of the Engenni Language.* Dallas TX: SIL.
Thomason, S. 2001. *Language Contact. An Introduction.* Edinburgh: Edinburgh University Press.
Thomason, S. & Everett, D. 2005. Pronoun borrowing. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society,* Vol 27 (2001) C. Chang, Houser Michael J., Y. Kim, D. Mortensen, M. Park-Doob & M. Toosarvandani (eds), 301–315. Berkeley CA: Berkeley Linguistics Society.
Thomason, S. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley CA: University of California Press.
Thompson, S., Longacre, R. & Hwang, S. 2007. Adverbial clauses. In Shopen (ed.), Vol. II: 237–300.
Tjia, J. 2007. *A Grammar of Mualang. An Ibanic Language of Western Kalimantan, Indonesia.* Utrecht: LOT.
Tonelli, A. 1926. *Grammatica e glossario della lingua degli ona-selknam della Terra del Fuoco.* Torino: Società Editrice Internazionale.
Tonhauser, J. 2007. Nominal tense? The meaning of Guaraní nominal temporal markers. *Language* 83(4): 831–869.

Topping, D. 1973. *Chamorro Reference Grammar.* Honolulu HI: University of Hawai'i Press.
Tosco, M. 2000. Is there an 'Ethiopian language area'? *Anthropological Linguistics* 42(3): 329–365.
Traill, A. 1985. *Phonetic and Phonological Studies of !Xóõ Bushman.* Hamburg: Helmut Buske.
Traill, A. 1994. *A !Xóõ Dictionary.* Köln: Rüdiger Köppe.
Trask, R. 1998. The typological position of Basque. Then and now. *Language Sciences* 20(3): 313–324.
Traugott, E. C. 2001. Legitimate counterexamples to unidirectionality. Paper presented at Freiburg University, October 17th 2001. <http://www.stanford.edu/~traugott/papers/Freiburg.Unidirect.pdf>
Trosterud, T. & Uibo, H. 2005. Consonant gradation in Estonian and Sámi. Two-level solution. In In*quiries into Words, Contsraints and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday,* A. Arppe, L. Carlson, K. Lindén, J. Piitulainen, M. Suominen, M. Vainio, H. Westerlund & A. Yli-Jyrä (eds), 136–150. Stanford CA: CSLI.
Trudgill, P. 2011. *Sociolinguistic Typology. The Social Determinants of Linguistic Structure and Complexity.* Oxford: OUP.
Tryon, D. 2002. Buma. In Lynch et al. (eds), 573–586.
Turner, B. 1986. *A Teaching Grammar of the Manam Language.* Ukarumpa: SIL.
Urbanová, L. 2007. Phatic communion and small talk in fictional dialogues. In *Functional Perspectives on Grammar and Discourse. In Honour of Angela Downing,* C. S. Butler, R. Hidalgo Downing & J. Lavid (eds), 349–58. Amsterdam: John Benjamins.
Ussishkin, A. 2006. Semitic morphology. Root-based or word-based? *Morphology* 16: 37–40.
Utas, B. 2009. Iranian languages as gardens. *Iranian Journal of Applied Language Studies* 1(1): 142–160.
Uyechi, L. 1996. The Navajo third person alternation and the pronoun incorporation analysis. In *Athabaskan Language Studies. Essays in Honor of Robert W. Young* E, Jelinek, S. Midgette, K. Rice & L. Saxon (eds), 123–136. Albuquerque NM: University of New Mexico Press.
Valli, C., Lucas, C. & Mulrooney, K. 2005. *Linguistics of American Sign Language. An Introduction,* 4th edn. Washington DC: Gallaudet University Press.
van den Berg, H. 2004. Coordinating constructions in Daghestanian languages. In *Coordinating Constructions,* M. Haspelmath (ed.), 197–226. Amsterdam: John Benjamins.
van der Auwera, J. (ed.). 1998. *Adverbial Constructions in the Languages of Europe.* Berlin: de Gruyter.
van der Auwera, J. & Ammann, A. 2011a. Overlap between situational and epistemic modal marking. In Dryer & Haspelmath (eds), Ch. 76. <http://wals.info/chapter/76> (11 May 2011).
van der Auwera, J. & Ammann, A. 2011b. Situational possibility. In Dryer & Haspelmath (eds), Ch. 74. <http://wals.info/chapter/74> (11 May 2011).
van der Auwera, J. & Lejeune, L. 2011a. The morphological imperative. In Dryer & Haspelmath (eds), Ch. 70. <http://wals.info/chapter/70> (11 May 2011).
van der Auwera, J. & Lejeune, L. 2011b. The prohibitive. In Dryer & Haspelmath (eds), Ch. 71. <http://wals.info/chapter/71> (11 May 2011).
van der Hulst, H. 1993. Units in the analysis of signs. *Phonology* 10: 209–241.
van der Kooij, E. 2002. *Phonological Categories in Sign Language of the Netherlands. The Role of Phonetic Implementation and Iconicity.* Utrecht: LOT.
van der Voort, H. 2013. Eskimo Pidgin. In Michaelis et al. (eds), Vol. III.
van Driem, G. 1987. *A Grammar of Limbu.* Berlin: Mouton de Gruyter.
van Driem, G. 1993. *A Grammar of Dumi.* Berlin: Mouton de Gruyter.
van Gelderen, E. 2006. *A History of the English Language.* Amsterdam: John Benjamins.
van Gijn, I. 2004. *The Quest for Syntactic Dependency. Sentential Complementation in Sign Language of the Netherlands.* Utrecht: LOT.

Van Valin, R. 1977. Aspects of Lakhota Syntax. PhD dissertation, University of California at Berkeley.
Van Valin, R. 2005. *Exploring the Syntax-Semantics Interface.* Cambridge: CUP.
Velupillai, V. 2003. *Hawai'i Creole English. A Typological Analysis of the Tense-Mood-Aspect System.* Basingstoke: Palgrave Macmillan.
Velupillai, V. & Hentschel, E. 2009. Tempus. In *Deutsche Morphologie,* E. Hentschel & P. M. Vogel (eds), 425–440. Berlin: Walter de Gruyter.
Veselinova, L. 2006. *Suppletion in Verb Paradigms. Bits and Pieces of the Puzzle.* Amsterdam: John Benjamins.
Vigário, M. 2003. *The Prosodic Word in European Portuguese.* Berlin: Mouton de Gruyter.
Voghel, A. 2005. Phonologically identical noun-verb pairs in Quebec Sign Language (LSQ). Form and context. *Toronto Working Papers in Linguistics* 25: 68–75.
Volker, C. 1994. Nalik Grammar (New Ireland, Papua New Guinea). PhD dissertation, University of Hawai'i.
von der Gabelentz, G. 1901[1891]. *Die Sprachwissenschaft. Ihre Aufgaben, Methoden und bisherigen Ergebnisse.* Leipzig: T. O. Weigel.
von Schlegel, F. 1808. *Über die Sprache und Weisheit der Indier.* Heidelberg: Mohr & Zimmer.
Wales, K. 1996. *Personal Pronouns in Present-Day English.* Cambridge: CUP.
Wali, K. & Koul, O. 1997. *Kashmiri. A Cognitive-Descriptive Grammar.* London: Routledge.
Wallace, S. 1983. Pronouns in contact. In *Essays in Honor of Charles F. Hockett,* F. B. Agard (ed.), 573–589. Leiden: Brill.
Wash, S. 2001. Adverbial Clauses in Barbareno Chumash Narrative Discourse (California, John Peabody Harrington). PhD dissertation, University of California, Santa Barbara.
Watson, S. 1976. The Pacoh noun phrase. *Mon-Khmer Studies* 5: 219–231.
Watters, D. 2002. *A Grammar of Kham.* Cambridge: CUP.
Watters, J. 1988. Topics in Tepehua Grammar. PhD dissertation, University of California, Berkeley.
Watts, R. 2003. *Politeness.* Cambridge: CUP.
Welmers, W. & Welmers, B. 1969. Noun modifiers in Igbo. *International Journal of American Linguistics* 35(4): 315–322.
West, B. 1980. *Gramatica popular del tucano.* Bogotá: Instituto Lingüístico de Verano.
Whaley, L. 1997. *Introduction to Typology. The Unity and Diversity of Language.* Thousand Oaks CA: Sage
Wichmann, S. & Wohlgemuth, J. 2008. Loan verbs in a typological perspective. In *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Rrocesses,* T. Stolz, D. Bakker & R. Salas Palomo (eds), 89–122. Berlin: de Gruyter.
Wiesemann, U. 1986. The pronoun systems of some Je and Macro-Je languages. In Wiesemann (ed.), 359–380.
Wiesemann, U. (ed.). 1986. *Pronominal Systems.* Tübingen: Narr.
Willett, M. 2003. A Grammatical Sketch of Nxa'Amxcin (Moses-Columbia Salish) (Washington). PhD dissertation, University of Victoria. Canada.
Wilson, J. 1996. A Phonological Grammar of Kuche (Nigeria). MA thesis, The University of Texas at Arlington.
Wolfart, H. 1973. *Plains Cree. A Grammatical Study.* Philadelphia PA: American Philosophical Society.
Wolff, H. 1993. *Referenzgrammatik des Hausa. Zur Begleitung des Fremdsprachenunterrichts und zur Einführung in das Selbststudium,* 2nd edn. Hamburg: LIT Verlag.
Woodward, J. 1978. Historical bases of American Sign Language. In *Understanding Language through Sign Language Research,* P. A. Siple (ed.), 333–348. New York NY: Academic Press.

Wright, D. 1995. The Verbal System of Logo (Logoti) (Sudan). MA thesis, The University of Texas at Arlington.
Wright, J. 1999. Nuer Noun Morphology. MA thesis, University of New York at Buffalo.
Yadav, R. 1996. *A Reference Grammar of Maithili.* Berlin: Mouton de Gruyter.
Yakpo, K. 2009. *A Grammar of Pichi.* Berlin: Isimu Media.
Yakpo, K. 2013. Pichi. In Michaelis et al. (eds), Vol. I.
Yip, M. 2002. *Tone.* Cambridge: CUP.
Yule, G. 1996. *Pragmatics.* Oxford: OUP.
Zaborski, A. 1991. Ethiopian language subareas. In *Unwritten Testimonies of the African Past. Proceedings of the International Symposium held in Ojrzanów n. Warsaw on 07–08 November 1989,* S. Pilaszewicz & E. Rzewuski (eds), 123–134. Warsaw: Warsaw University Press.
Zeitoun, E., Huang, L., Yeh, M., Chang, A. & Wu, J. 1996. The temporal, aspectual, and modal systems of some Formosan languages. A typological perspective. *Oceanic Linguistics* 35(1): 21–56.
Zeshan, U. 2002. Towards a notion of 'word' in Sign Languages. In *Word. A Cross-Linguistic Typology,* R. M. W. Dixon & A. Aikhenvald (eds), 153–79. Cambridge: CUP.
Zeshan, U. 2003a. Aspects of Türk İşaret Dili (Turkish Sign Language). *Sign Language & Linguistics* 6(1): 43–75.
Zeshan, U. 2003b. 'Classificatory' constructions in Indo-Pakistani Sign Language. Grammaticalization and lexicalization processes. In *Perspectives on Classifier Constructions in Sign Language,* K. Emmorey (ed.), 113–141. Mahwah NJ: Lawrence Erlbaum Associates.
Zeshan, U. 2003c. Indo-Pakistani Sign Language grammar. A typological outline. *Sign Language Studies* 3(2): 157–212.
Zeshan, U. 2004a. Hand, head, and face. Negative constructions in Sign Languages. *Linguistic Typology* 8: 1–58.
Zeshan, U. 2004b. Interrogative constructions in signed languages. Crosslinguistic perspectives. *Language* 80(1): 7–39.
Zeshan, U. 2011a. Irregular negatives in Sign Languages. In Dryer & Haspelmath (eds), Ch. 139. <http://wals.info/chapter/139> (11 May 2011).
Zeshan, U. 2011b. Question particles in Sign Languages. In Dryer & Haspelmath (eds), Ch. 140. <http://wals.info/chapter/140> (11 May 2011).
Zeshan, U. 2011c. Sign Languages. In Dryer & Haspelmath (eds), Supplement 9. <http://wals.info/supplement/9> (11 May 2011).
Zilles, A. 2005. The development of a new pronoun. The linguistic and social embedding of *a gente* in Brazilian Portuguese. *Language Variation and Change* 17: 19–53.
Zúñiga, F. 2000. *Mapudungun.* München: Lincom.
Zwitserlood, I. 2003. *Classifying Hand Configurations in Netherlandse Gebarentaal (Sign Language of the Netherlands).* Utrecht: LOT.
Zylstra, C. 1991. A syntactic sketch of Alacatlatzala Mixtec. In *Studies in the Syntax of Mixtecan Languages,* Vol. III, C. H. Bradley & B. E. Hollenbach (eds), 1–178. Dallas TX: SIL.

Subject index