GRAMMATICALIZATION FROM A TYPOLOGICAL PERSPECTIVE

Edited by

HEIKO NARROG
and BERND HEINE

Grammaticalization from a Typological Perspective

OXFORD STUDIES IN DIACHRONIC AND HISTORICAL LINGUISTICS

GENERAL EDITORS

Adam Ledgeway and Ian Roberts, *University of Cambridge*

ADVISORY EDITORS

Cynthia Allen, *Australian National University*; Ricardo Bermúdez-Otero, *University of Manchester*; Theresa Biberauer, *University of Cambridge*; Charlotte Galves, *University of Campinas*; Geoff Horrocks, *University of Cambridge*; Paul Kiparsky, *Stanford University*; Anthony Kroch, *University of Pennsylvania*; David Lightfoot, *Georgetown University*; Giuseppe Longobardi, *University of York*; George Walkden, *University of Konstanz;* David Willis, *University of Cambridge*

Grammaticalization from a Typological Perspective

Edited by
HEIKO NARROG
and
BERND HEINE

OXFORD
UNIVERSITY PRESS

Contents

Series preface

Modern diachronic linguistics has important contacts with other subdisciplines, notably first-language acquisition, learnability theory, computational linguistics, sociolinguistics, and the traditional philological study of texts. It is now recognized in the wider field that diachronic linguistics can make a novel contribution to linguistic theory, to historical linguistics, and arguably to cognitive science more widely.

This series provides a forum for work in both diachronic and historical linguistics, including work on change in grammar, sound, and meaning within and across languages; synchronic studies of languages in the past; and descriptive histories of one or more languages. It is intended to reflect and encourage the links between these subjects and fields such as those mentioned above.

The goal of the series is to publish high-quality monographs and collections of papers in diachronic linguistics generally, i.e. studies focusing on change in linguistic structure, and/or change in grammars, which are also intended to make a contribution to linguistic theory, by developing and adopting a current theoretical model, by raising wider questions concerning the nature of language change, or by developing theoretical connections with other areas of linguistics and cognitive science as listed above. There is no bias towards a particular language or language family, or towards a particular theoretical framework; work in all theoretical frameworks, and work based on the descriptive tradition of language typology, as well as quantitatively based work using theoretical ideas, also feature in the series.

Adam Ledgeway and Ian Roberts

University of Cambridge

Preface

This volume has emerged out of a symposium with the title 'Grammaticalization in Japanese and Across Languages', held at the National Institute for Japanese Language and Linguistics, Tokyo, 3rd to 5th July 2015. We wish to thank the head of the National Institute, Professor Taro Kageyama, and the head of the sponsoring department, Professor Prashant Pardeshi, for their generous and kind support that enabled this occasion.

Proceeding with the book project, we received valuable advice from Oxford University Press and two anonymous peer reviewers. The feedback that we received, and the circumstances of some the original contributors, led to a reshaping. We are very grateful for the feedback, and very much obliged to those colleagues who had the patience to stick with the project, and finally to those who agreed to newly join, at a slight disadvantage compared with the 'first-generation' collaborators, in terms of familiarity with the task, and time pressure.

List of abbreviations

1	1st person
2	2nd person
3	3rd person
>	interclause correferentiality relation
A	agent/subject of transitive verb
ABL	ablative
ABS	absolutive
ACC	accusative
ACT	actor
ADD	additive
ADES	adessive
ADJ	adjective
ADN	adnominal
ADV	adverbial
AGT	agentive
ALIEN	alienable
ALL	allative
ANA	anaphoric
ANT	anterior tense
AOR	aorist
APPL	applicative
APUD	'near' localization
ASP	Aspect
ASS	assertive
ASSC	associative
ATTR	attributive
AUX	auxiliary
BEN	benefactive
BOU	boulomaic modality (intention)
BU	buffer element
CAU	causative
CFUG	centrifugal directional
CL	nominal class marker
CLF	classifier

CNTR	contrastive
CNV	converb
COM	comitative
COMP	comparative
COND	conditional
CONJ	conjunction
CON	conative
CONN	connective
CONT	continuative
COORD	coordination
COP	Copula
CPL	complementizer
DAT	dative
DEC	declarative
DEF	definite
DEIC	deictic
DEM	demonstrative
DEP	dependent
DES	desiderative
DETR	detransitive
DIR	directional
DIST	distal demonstrative
DM	discourse marker
DU	dual
DUR	durative
ELAT	elative
EMPH	emphatic
EMSEA	East and Mainland South East Asia
EP	epenthetic
ERG	ergative
ESS	essive
EVID	evidential
EXHORT	exhortative
EXCL	exclusive
EXP	experiential
EZ	ezafe (linking) particle
F, f	feminine
FOC	focus
FREQ	frequentative

FUT	future
GEN	genitive
GER	gerund(ive)
HAB	habitual
HON	honorific
HORT	hortative
HRSY	hearsay
HUM	human
IMM	immediate future
IMP	imperative
IMPF	imperfect
IMPV	imperfective
IN	'inside' localization
INC	inceptive
INC	incompletive
INCH	inchoative
INCL	inclusive
INCON	inconsequential
IND	indicative (mood)
INDF	indefinite
INESS	inessive
INF	infinitive
INJ	interjection
INS	instrumental
INT	intensive
INTR	intransitive
IO	indirect object
IRR	irrealis
JUSS	jussive
LAT	lative
LK	linker
LOC	locative
LV	loan verb
M, m	masculine
MAL	malefactive
MID	middle
MSD	masdar
n	neuter
N	noun

N1	noun class 1
NAR	narrative
NEG	negation
NF, nf	non-feminine
NFIL	non-final
NFIN	non-finite
NFUT	non-future
NI	new information
NMLZ	nominalizer
NOM	nominative
NON.POSS	non possessed
NON.PROX	non proximal to the addressee
NP	Noun Phrase
NPS	non-past tense
NRL	non-relational noun prefix
NSG	non-singular
NSIT	new situation
NSPC	non-specific article
NVIS	non-visible
OBJ	object
OBL	oblique
OBLIG	obligation
OBV	obviative
ONOM	onomatopoeia
ORD	ordinal
P	patient/direct object of transitive verb
PASS	passive
PAST.VIS.INTER	visual past interrogative
PE	previous event
PEOc	Proto-Eastern Oceanic
PERM	permissive
PERS	personal
PERT	pertensive
PFV	perfective
PL, pl	plural
PN	proper noun
PNP	Preposition-Noun-Preposition
POc	Proto-Oceanic
POSS	possessive

POSS.LINK	Possessive Linker
POST	'behind' localization
POT	potential
PP	past participle
PPP	past/perfective passive participle
PR	possessor
PRET	preterite
PREX	prefix
PRF	perfect
PROG	progressive
PROP	proprietive
PROX	proximate
PRS	present
PRS.NONVIS	present non-visual
PRS.VIS	present visual
PRV	preverb
Ps	person
PST	past
PST.IMP	past imperfect
PTC	particle
PTCP	participle
PURP	purposive
Q	question/interrogative
QUAL	qualitative predication
QUO	quotative
R	Russian loan
RDP	reduplication
RE	refactive
REAL	realis
REL	relativizer
REM.P	remote past
REM.P.REP	remote past reported
REM.P.VIS	remote past visual
REP	repetitive
RESTR	restrictive particle
RETR	retrospective
RFL	reflexive
RL	relational noun prefix
RPR	reportative

RSN	reason
S	single argument of an intransitive predicate
S/A	S/A indexing suffix
SAE	Standard Average European
SBD	subordinator
SBJ	subjunctive mood
SBZ	substantivizer
SE	simultaneous event
SENS	sensory
SEQ	sequential
SFP	sentence-final particle
SG, sg	singular
SPC	specific article
SPEC	specific
SS	same subject
STAT	stative
SUB	'under' localization
SUBJ	subject
SUBO	subordinate
SUCC	successive aspect
SUR	surpass
SVC	serial verb construction
TEMP	temporal
TERM	terminative
TOP	topic
TOP.NON.A/S	topical non-subject
TR	transitive
U	undergoer
V	Verb
VB	verbalizer
VENT	ventive
VERIF	verificative
VOC	vocative

The contributors

ALEXANDRA Y. AIKHENVALD is Distinguished Professor, Australian Laureate Fellow, and Director of the Language and Culture Research Centre at James Cook University. She is a major authority on languages of the Arawak family, from northern Amazonia, and has written grammars of Bare (1995) and Warekena (1998), plus *A Grammar of Tariana, from northwest Amazonia* (Cambridge University Press, 2003), in addition to essays on various typological and areal features of South American languages. Her other major publications, with Oxford University Press, include *Classifiers: A Typology of Noun Categorization Devices* (2000), *Language Contact in Amazonia* (2002), *Evidentiality* (2004), *The Manambu Language from East Sepik, Papua New Guinea* (2008), *Imperatives and Commands* (2010), *Languages of the Amazon* (2012), *The Art of Grammar* (2015), *How Gender Shapes the World* (2016), and *Serial Verbs* (2018).

UMBERTO ANSALDO is Professor in Linguistics and Head of the School of Literature, Arts and Media at The University of Sydney. He specializes in language contact and typology of the Asian region. He is the author of *Languages in Contact: Ecology and Evolution in Asia* (Cambridge University Press, 2009) and the co-editor in chief of the journal *Language Ecology* (John Benjamins).

PETER ARKADIEV holds a PhD in theoretical, typological, and comparative linguistics from the Russian State University for the Humanities. Currently he is a senior researcher at the Institute of Slavic Studies of the Russian Academy of Sciences. His fields of interest include language typology and areal linguistics, case and alignment systems, tense-aspect, and Baltic and Caucasian languages. He has co-edited *Borrowed Morphology* (with Francesco Gardani and Nino Amiridze, de Gruyter, 2014) and *Contemporary Approaches to Baltic Linguistics* (with Axel Holvoet and Björn Wiemer, de Gruyter, 2015).

WALTER BISANG studied General Linguistics, Chinese, and Georgian at the University of Zürich. He has been Professor of General and Comparative Linguistics at the University of Mainz (Germany) since 1992, and Director of the Collaborative Research Center on 'Cultural and Linguistic Contact' from 1999 to 2008. His research interests include grammaticalization, linguistic typology, and language contact/areal typology. His languages of interest are: East and mainland Southeast Asian languages, Caucasian languages (Georgian), Austronesian languages (Bahasa Indonesia, Tagalog, Yabêm, Paiwan), and Yoruba (together with Remi Sonaiya).

ALEXANDER R. COUPE is Associate Professor of Linguistics and Multilingual Studies at Nanyang Technological University, Singapore. He is the author of *A Grammar of Mongsen Ao* (2007) and has a major research focus on the documentation and analysis of the Tibeto-Burman, Indo-Aryan, and Austroasiatic languages of Northeast India, having worked on the languages of this region for over two decades. The output of this research feeds his theoretical interests in linguistic typology, grammaticalization theory, and the development of complexity in the grammars of the world's languages.

ÉVA Á. CSATÓ is Professor Emeritus in Turkic Languages at the Department of Linguistics and Philology, Uppsala University, Sweden. She has made contributions to the typology of Turkic languages as well as the documentation of less-studied and endangered Turkic varieties. Together with Lars Johanson she edited the volume *The Turkic Languages* (1998). She is a member of the editorial boards of the journals *Turkic Languages* (Harrassowitz).

ÖSTEN DAHL is Professor Emeritus of General Linguistics at Stockholm University, Sweden. He obtained his academic training at the universities of Gothenburg, Uppsala, and Leningrad (St Petersburg), and was active at the University of Gothenburg for ten years before moving to Stockholm in 1980. In recent years, his research has mainly been typologically oriented with a strong interest in diachronic approaches to grammar. He has published the monographs *Tense and Aspect Systems* (1985), *The Growth and Maintenance of Linguistic Complexity* (2004), and *Grammaticalization in the North: Noun Phrase Morphosyntax in Scandinavian Vernaculars* (2015).

MOHSSEN ESSEESY, PhD, is an Associate Professor of Arabic and International Affairs and the Chair of the Department of Classical and Near Eastern Languages and Civilizations at George Washington University in Washington, DC. His publications include *Grammaticalization of Arabic Prepositions and Subordinators: A Corpus-Based Study* (Brill, 2010); *Contemporary Business Arabic* (Georgetown University Press, forthcoming); Arabic for Specific Purposes in *Handbook for Arabic Teaching Professionals in the 21st Century*, vol. 2 (Routledge, 2017). Dr Esseesy is a member of the Editorial Board of the series Routledge Studies in Language and Identity, and was an elected member the Executive Board of the American Association of Teachers of Arabic, 2004–2006.

GEOFFREY HAIG is Professor of General Linguistics at the Institute for Oriental Studies at the University of Bamberg. He received his PhD in 1997 at the University of Kiel, and has since worked as a full-time researcher and lecturer in linguistics, focusing on corpus-based approaches to language typology, the historical syntax of Iranian, areal linguistics (Western Asia), and Kurdish (see https://www.uni-bamberg.de/aspra/mitarbeiter/prof-dr-geoffrey-haig/).

MARTIN HASPELMATH is senior scientist at the Max Planck Institute for the Science of Human History (Jena) and Honorary Professor at Leipzig University. Between 1998 and 2015, he was a member of the linguistics department of the Max Planck Institute for Evolutionary Anthropology (Leipzig). His research interests are primarily in the area of broadly comparative and diachronic morphosyntax (*Indefinite Pronouns*, 1997; *Understanding Morphology*, 2002) and in language contact (*Loanwords in the World's Languages*, co-edited with Uri Tadmor, 2009). He is one of the editors of Oxford University Press's *World Atlas of Language Structures* (2005), and of the *Atlas of Pidgin and Creole Language Structures* (2013).

BERND HEINE is Emeritus Professor at the Institute of African Studies (Institut für Afrikanistik), University of Cologne, Germany. He is presently Yunshan Chair Professor at Guangdong University of Foreign Studies, China. His 33 books include *Possession: Cognitive Sources, Forces, and Grammaticalization* (Cambridge University Press, 1997); *Cognitive Foundations of Grammar* (Oxford University Press, 1997); with Tania Kuteva, *World Lexicon of Grammaticalization* (Cambridge University Press, 2002); *Language Contact and Grammatical Change* (Cambridge University Press, 2005). He has held visiting professorships in Europe, Eastern

Asia (Japan, South Korea, China), Australia (LaTrobe University, Melbourne), Africa (University of Nairobi, University of Cape Town), North America (University of New Mexico, Dartmouth College), and South America (Universidade Federal Fluminense, Rio de Janeiro). He has been a fellow of the Center for Advanced Study in the Behavioral Sciences, Stanford, USA (1999–2000), the Netherlands Institute for Advanced Study (NIAS), Wassenaar (2005–2006), and Tokyo University of Foreign Studies (2008–2009). His present main research areas are discourse grammar and grammaticalization theory.

LARS JOHANSON is a linguist and Turcologist. For many years he has been Professor of Turcology at the Department of Oriental Studies of the University of Mainz, Germany, and a Senior Lecturer in Turkic Languages at Uppsala University, Sweden. He has been instrumental in transforming the field of Turcology, which was traditionally more philologically oriented, into a linguistic discipline. Apart from his contributions to Turcology, Lars Johanson has made a number of pioneering contributions to general linguistics and language typology, in particular to the typology of viewpoint aspect systems and the theory of language contact. He is the editor of the journal *Turkic Languages* (Harrassowitz) and of the monograph series *Turcologica* (Harrassowitz).

MARIAN KLAMER is Professor of Austronesian and Papuan Linguistics at Leiden University. Over the last two decades she has led research projects describing and documenting Austronesian and Papuan minority languages in Eastern Indonesia, comparing their typological characteristics and reconstructing their history. Her publications include grammars on Kambera (1998), Teiwa (2010), and Alorese (2011), as well as several edited volumes, and over fifty articles on a wide range of topics, including language contact and grammaticalization in Indonesia. She is currently leading the NWO-VICI project 'Reconstructing the past through languages of the present: the Lesser Sunda Islands' (2014–2019).

TIMUR MAISAK holds a PhD in theoretical, typological, and comparative linguistics from Lomonosov Moscow State University (2002). Currently he is a senior researcher at the Institute of Linguistics of the Russian Academy of Sciences. His fields of interest include language typology and language documentation, grammaticalization theory, tense and aspect systems, and East Caucasian (Nakh-Daghestanian) languages. He has co-edited *Tense, Mood, Aspect and Finiteness in East Caucasian Languages* (with Gilles Authier, Brockmeyer, 2011) and *The Semantics of Verbal Categories in Nakh-Daghestanian Languages: Tense, Aspect, Evidentiality, Mood/Modality* (with Diana Forker, Brill, 2018).

JOHN H. MCWHORTER is Professor of Linguistics, Philosophy and Music History at Columbia University. He is the author of *The Missing Spanish Creoles*, *Language Interrupted*, *Defining Creole*, and *Linguistic Simplicity and Complexity*, as well as trade books such as *The Power of Babel*, *Our Magnificent Bastard Tongue*, *The Language Hoax*, and *Words on the Move*.

MARIANNE MITHUN is Professor of Linguistics at the University of California, Santa Barbara. Her work focuses on morphology, syntax, discourse, prosody, and their interrelations; language contact and language change; typology and universals; language documentation; and the linguistics of languages indigenous to North America and eastern Austronesia.

CLAIRE MOYSE-FAURIE is a French linguist from the Lacito-CNRS, Paris. Since 1976, she has done fieldwork on several Kanak languages (Drehu, Xârâcùù, Xârâgurè, West Uvean, and

Haméa) and on two Polynesian languages (East Futunan and East Uvean), elaborating dictionaries, grammars, and oral tradition texts. She is also involved in several typological projects (such as categorization, nominalization, reflexives and reciprocals, valency and verb classes, and spatial deixis), as well as in documentation projects on endangered languages.

Ilana Mushin is Associate Professor and Reader in Linguistics at the University of Queensland, Australia, specializing in pragmatics, typology, and the description of Australian languages. She is the author of *A Grammar of (Western) Garrwa* (2012, Mouton de Gruyter) and co-editor of *Discourse and Grammar in Australian Languages* (2008, Benjamins). Her research into Garrwa grammar has focused on the pragmatic underpinnings of grammatical development, especially pronouns systems and the development of second-position phenomena in Australian languages.

Heiko Narrog is Professor at Tohoku University, Japan. He received a PhD in Japanese studies from the Ruhr University Bochum in 1997, and a PhD in language studies from Tokyo University in 2002. His publications include *Modality in Japanese and the Layered Structure of Clause* (Benjamins, 2009), *Modality, Subjectivity, and Semantic Change: A Cross-Linguistic Perspective* (Oxford University Press, 2012), and *The Oxford Handbook of Grammaticalization* (Oxford University Press, 2011) and *The Oxford Handbook of Linguistic Analysis*, 2nd edn (Oxford University Press, 2015), both co-edited with Bernd Heine.

Seongha Rhee is Professor at Hankuk University of Foreign Studies, Korea. He received a PhD in Linguistics from the University of Texas at Austin in 1996. He served as President of the Linguistic Society of Korea, 2013–2014, and as President of the Discourse and Cognitive Linguistic Society of Korea, 2009–2011. He has published book chapters in *The Oxford Handbook of Grammaticalization* (Oxford University Press, 2011), *Rethinking Grammaticalization* (Benjamins, 2008), *Shared Grammaticalization* (Benjamins, 2013) among others, and has published research articles in journals including *Journal of Pragmatics, Language Sciences*, and *Journal of Historical Pragmatics*.

Hiram L. Smith received a PhD in Spanish Linguistics from Penn State University in 2014. He currently works as Assistant Professor of Spanish and Linguistics at Bucknell University, Lewisburg, Pennsylvania. His research focuses on language variation and change, with an emphasis on marginalized speech communities in Spanish-speaking and Anglophone America. Currently, he is researching change in Palenquero Creole, spoken in the Afro-Hispanic village of San Basilio de Palenque, Colombia. An integral part of his research philosophy is that whenever possible community members should take part in the scientific study of their own community, and that researchers should always put science at the service of the community.

Pui Yiu Szeto is a PhD candidate at the Department of Linguistics, University of Hong Kong. His research interests lie primarily in the field of language contact, particularly in contact-induced grammaticalization as well as the interplay between linguistic and social factors in the emergence and evolution of contact varieties.

John Whitman is a Professor in and Chair of the Department of Linguistics at Cornell University, New York. His linguistic specializations are East Asian linguistics, comparative syntax, language typology, and historical linguistics. He is currently editor of the journal *Korean Linguistics* (Benjamins). Recent publications in linguistics include a co-edited volume *Ryūkyū shogo to Kodai Nihongo: Nichiryū sogo no saiken ni mukete* (Ryūkyūan and premodern

Japanese: toward the reconstruction of Proto-Japanese-Ryūkyūan, 2016, with Yukinori Takubo and Tatsuya Hirako), the chapter 'Old Korean' in the Blackwell *Handbook of Korean Linguistics* (2015), and the chapter on 'Topic prominence' in the *Blackwell Companion to Syntax* (with Waltraud Paul, to appear).

Roberto Zariquiey holds a PhD in Linguistics from LaTrobe University, Melbourne, Australia. As his PhD thesis, he wrote a reference grammar of Kakataibo, which was granted an honourable mention in the prestigious Panini Award (from the Association for Linguistic Typology). This grammar is about to be published in the Grammar Library of Mouton de Gruyter (2018). He is currently an associate professor at the Pontificia Universidad Católica del Perú (PUCP), where he also is the Director of the Masters Program in Linguistics and the Co-Director (and founder) of the Digital Archive of Peruvian Languages. In recent years, he has been a visiting scholar in prestigious institutions such as the University of California at Berkeley (2014), the Max Planck Institute for Evolutionary Anthropology (MPI-EVA) (2014), and the Collegium de Lyon (2017).

1

Introduction
Typology and grammaticalization

HEIKO NARROG AND BERND HEINE

1.1 INTRODUCTION

The goal of this volume is to identify aspects of grammaticalization that correlate with typological features of languages. Previously, the hypothesis that certain criteria of grammaticalization may apply differently to different types of languages, and concomitantly language areas, was mainly raised with respect to Southeast and East Asian languages (Bisang 2004, 2008, 2011; Narrog and Ohori 2011). Our idea, then, was to pursue this hypothesis by casting the net wider and more systematically, and invite experts on different language areas to reflect on the relationship between language type and language area on the one hand and grammaticalization on the other.

To start out with the basics, we define typology and grammaticalization as follows:

(1) Linguistic typology 'concerns itself with the study of structural differences and similarities between languages. [...] [It] is the study and interpretation of linguistic or language types' (Velupillai 2012: 15).

(2) Grammaticalization concerns 'the way grammatical forms arise and develop through space and time' (Heine 2003: 575).

The process of grammaticalization can be divided into the four basic aspects listed in (3) (cf. Heine and Narrog 2010: 405).

(3) (a) extension (or context generalization): use in new contexts;
     (b) desemanticization: loss (or generalization) in meaning content;
     (c) decategorialization: loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms; and
     (d) erosion (or 'phonetic reduction'): loss in phonetic substance.

In our view, the pragmatic-semantic processes (3a) and (3b) are essential for grammaticalization. That is, we consider semantic change as the core of grammaticalization,

as will also be argued in Heine (Chapter 2 this volume). The structural processes given in (3c,d) may or may not follow, and if they do, by hypothesis they do so in the order of the list in (3).

If there is typological variation in grammaticalization, it may in principle affect all aspects of grammaticalization. There is no reason to exclude some aspect a priori. However, if our hypothesis is correct that semantic change is the essence of grammaticalization and other aspects follow in the order given above, we can hypothesize that the order in which typological variation in grammaticalization occurs is in reverse: (d) > (c) > (b) > (a). Accordingly, formal reductive change should be most susceptible to variation and semantic/functional change least susceptible.

From the literature on grammaticalization and typology cited in the first paragraph of this section, one may infer that it is primarily extant typological features that in some way influence grammaticalization. But interaction between grammaticalization and typology does not have to go in one direction. As argued in Narrog (2017a), the two directions of influence in (4) and (5) are not only hypothetically possible but have also been empirically observed.

(4)   Typological features influence aspects of grammaticalization.

(5)   Grammaticalization motivates structural features that can be typologized.

Furthermore, typological features interacting with grammaticalization have been identified with certain language areas or groups of languages rather than individual languages. This is due to the fact that structural features are often shared between areally adjacent languages, like the East and mainland Southeast Asian languages, sometimes across genetic boundaries. Therefore, while the target of our project is the interaction between typological features of languages and grammaticalization, areal groups of languages that share structural features have been chosen as the subject of most of the studies in this project.

Sections 1.2 and 1.3 elaborate on (4) and (5) above, and refer to chapters in this volume that show these influences at work. Section 1.3 also contains a short conclusion. Finally, section 1.4 provides a conclusion and a brief overview of the structure of this volume.

1.2 TYPOLOGICAL FEATURES INFLUENCE GRAMMATICALIZATION

It is a long-standing question what ultimately drives, or motivates, grammaticalization. Haspelmath (1999, Chapter 6 this volume), for example, has been arguing that inflation on the one hand and extravagance on the other hand are the ultimate causes of a perpetual cycle of grammaticalization. For Hawkins (2011), efficiency and ease of processing are the causes of language change and grammaticalization, as he considers grammars as the 'conventionalizations of the same processing mechanisms that psychologists find evidence for in experimental and corpus data' (p. 225).When it comes to the shape taken by new grammaticalizations, or the paths through which

they grammaticalize, it is only reasonable to assume that extant language structures, including typologically relevant structures, may play a role, even if it is difficult (if possible at all) to demonstrate a clear cause–effect relationship between older structures and the form of emerging structures, since 1:1 replication of an older structure would be a rarity.

One also has to keep in mind that the definition of grammaticalization adopted also determines the extent to which grammaticalization can be analysed as being influenced by typological features. Concretely, the more grammaticalization is reduced to a universal essence, the less room there is for variation. Himmelmann (2005: 82), for example, claims that 'a grammaticisation process can be defined as a process of context expansion' on the three levels of (a) host class formation/expansion, (b) syntactic context expansion, and (c) semantic-pragmatic context expansion. This definition intentionally excludes typological factors from grammaticalization, which are viewed as epiphenomenal. Himmelmann (p. 83) explicitly states:

(6)　　The above definition of grammaticisation [...] differs from previous definitions of grammaticisation in [...] singling out context expansion in general and semantic-pragmatic context expansion in particular as the major defining feature of grammaticisation. All the other phenomena which are often observed in grammaticisation processes and which are considered criterial in other definitions [...] do not occur in this definition [...] [E]rosion and fusion here are considered epiphenomena. Their occurrence depends on at least two factors. For one, they depend on the overall typological profile of a given language (e.g. in isolating languages the potential for fusion is generally very limited). For another, the construction type also appears to play a major role.

In generative approaches to grammaticalization as well, grammaticalization tends to be reduced to an essence that is universal and not amenable to typological influences. Thus, for example, Roberts and Roussou (1999: 1011) regard grammaticalization 'as the diachronic development of lexical heads into functional heads', and van Gelderen (2004) conceptualizes grammaticalization as change from lower head to higher head ('Late Merge') or from Spec to Head. There might be some typological variation in the heads that are the sources and the targets of change, but the essence will remain unaffected. In this book, however, we are interested in the 'full package' of grammaticalization, including phonological and morphological changes. This full package is the subject of the remainder of this section.

Phonological and morphological typological features may be a guiding factor for grammaticalization processes in several ways. First, it has been claimed that grammaticalization in tonal, isolating languages, like the Sinitic languages, does not lead to reduction of syllables. Ansaldo and Lim (2004: 345) suggest that '[s]trongly isolating languages typically do not allow yesterday's syntax to become today's morphology [...] syllable boundaries are discrete and phonotactic constraints rule out reduced syllables of the kind observed elsewhere, the material available for reduction is not easily found at the morphological level' (cf. also Bisang 2004, 2010). In the same vein, Ansaldo, Bisang, and Szeto (Chapter 11 this volume) claim that 'elaboration of morphological structure only happens in a certain type of languages'. That is, 'the formal aspects of canonical

grammaticalization do not happen in [East and mainland Southeast Asian] languages.' Narrog, Rhee, and Whitman (Chapter 9 this volume; also Narrog and Ohori 2011) point out that, in contrast to the situation in the 'isolating' languages of East and mainland Southeast Asia, morphological parameters of grammaticalization seem to apply particularly well to Northeast Asian languages like Korean and Japanese, and probably to the so-called Transeurasian languages in general, which tend to be agglutinating. Mithun (Chapter 15 this volume) emphasizes the relatively high degree of morphological complexity as a product of extensive grammaticalization in North American languages. While the most common path leads from grammaticalization to univerbation, in North American languages we also often find the reverse, namely univerbation preceding grammaticalization. This is a source of a degree of polysynthesis that is not found in many language areas of the world. On the other hand, the fact that morphological complexity is the result of grammaticalization also points to the fact that there is a two-way relationship between grammaticalization and typological features— that is, typological features do not unilaterally determine features of grammaticalization.

In the case of the languages described by Bisang, structural typological features seem to perpetuate themselves horizontally, so to speak, across languages and vertically across the history of individual languages. Very generally, we may assume that, unless there is some disruption, the typological morphological tendencies of a language result in the extant morphological types as target structures: for example, in languages with agglutinative morphology, grammaticalization is more likely to lead to affixation than in isolating languages.

Dahl (Chapter 5 this volume), while in general not agreeing with Ansaldo et al.'s view of different rates of grammaticalization in East and mainland Southeast Asian versus European languages, nevertheless agrees that 'the likelihood for a certain grammaticalization process to appear is at least to some extent dependent on structural properties of the language'. Concretely, the inflectional endings on verbs and nouns that modern European languages have still preserved (but to a considerable extent already lost) are likely to be a remnant of grammaticalization in head-final proto-Indo-European.

But there are also synchronic semantic and syntactic features of languages that may guide grammaticalization. First, as discussed in more detail in section 1.3, in head-final languages grammaticalization is more likely to lead to bound morphemes than in head-initial languages because of the tendency for postposed rather than preposed morphemes to become bound. This theme is echoed by Dahl (Chapter 5 this volume) and by Narrog, Rhee, and Whitman (Chapter 9), as already seen above.

Second, Bisang (2004, 2008, 2011) claims a number of constraints on grammaticalization in Sinitic languages, which stand representatively for the East and mainland Southeast Asian languages. Besides the small extent of changes in phonology and morphology, these include (i) lack of obligatoriness of grammatical categories; i.e. lack of grammatical paradigms in general, (ii) lack of clearly determined semantic domains, (iii) existence of rigid word-order patterns within which lexical items grammaticalize, and pervasiveness of inference, which enables language users to encode and decode the function of a specific item in a specific semantic context, (iv) no grammaticalization chains (i.e. continuous grammaticalization from one category to the next). In other

words, 'widespread polyfunctionality undermines the semantic dimension of canonical grammaticalization in [East and mainland Southeast Asian] languages'. Accordingly, none of Lehmann's (2002) commonly cited six parameters of grammaticalization applies except syntagmatic variability, nor do other traditional models of grammaticalization. But even this parameter, according to Ansaldo, Bisang, and Szeto (Chapter 11 this volume) appears to be dubious: 'An aspect of grammaticalization in this area may be the loss of autonomy, or constructionalization, but even this is undermined by polyfunctionality and lack of obligatory marking.'

Esseesy (Chapter 3 this volume) does not recognize such kinds of constraints of grammaticalization in Arabic languages. He writes that 'there appears to be no typological limit found on the evolution of meaning and form in Semitic of the type described in Bisang's (2011) study'. Esseesy highlights the path of one specific gram, *fī>f-* 'in, at' that underwent rampant polygrammaticalization. It may go without saying that there is also a lot of intra-language variation, or variation within languages of one group with respect to degrees of grammaticalization. This is amply shown not only by Esseesy (Chapter 3) but also by Haig (Chapter 4), Mushin (Chapter 13), Moyse-Faurie (Chapter 14), and Arkadiev and Maisak (Chapter 7). Thus, Arkadiev and Maisak demonstrate that 'grammaticalizing constructions in North Caucasian languages display various degrees of integration, ranging from highly autonomous auxiliaries to those partly or totally fused with lexical verbs, up to the extent of becoming affixes'.

This kind of replication of existing structures through grammaticalization may apply not only to morphology (and syntax) but also to grammatical functions. Ideally speaking, every language would have just one exponent for every cross-linguistically available grammatical category. In reality, in any given language, expressions for specific categories may be grammaticalized en masse, while other categories are not grammaticalized at all. These categories may be expressed then indirectly through other categories, lexically, or not at all. As a result, a certain set of categories is grammaticalized over and over in waves or cycles. Those dominant categories may further impact the way other categories are grammaticalized.

In the area of tense-aspect-mood (TAM), Bhat (1999) suggests that languages tend to be either tense-, aspect-, or mood-prominent, and the categories which are not overtly expressed are often indirectly expressed through the well-grammaticalized ones. In terms of grammaticalization, we can surmise that languages tend to maintain this typological profile by repeatedly grammaticalizing their preferred category rather than grammaticalizing the 'neglected' ones. Bhat (1999: 182) concretely suggests different paths of grammaticalization: '[L]anguages that give greater prominence to aspect than to tense develop a perfective form from an earlier perfect construction and an imperfective form from an earlier progressive construction, whereas languages that give greater prominence to tense than to aspect develop past and present forms directly from their perfect and progressive constructions respectively.' Furthermore,

aspect-prominent and mood-prominent languages show distinct tendencies of change when they develop temporal distinctions. In the case of aspect-prominent languages, we generally find a two-way past/non-past distinction or a three-way past-present-future distinction

developing from an earlier perfective-imperfective distinction [...]. In the case of mood-prominent languages, on the other hand, the general tendency is to develop primarily a future/non-future distinction.   (Bhat 1999: 182–3)

Chafe (2000 presents a related phenomenon with his idea of 'florescence': 'Like forests, languages may develop toward a climax stage where particular combinations of features, like plant communities, may flourish to define a particular language type. I think it is useful to think in terms of the *florescence* of linguistic features in this sense— the flowering of features that come to dominate the form a language takes' (p. 39). Some features of Iroquoian languages serve as examples. Firstly, especially Northern Iroquoian languages can have an elaborate inventory of up to 60–70 pronominal prefixes. Besides singular, plural, and dual number distinctions in the first and second persons of both agents and patients, Cherokee has even an inclusive/exclusive distinction and gender distinctions with third persons (cf. Chafe 2002: 41). This elaborate pronominal prefix system is described (reconstructed) by Chafe as the result of not one but successive waves of grammaticalizations. Secondly, Iroquoian languages are characterized by noun incorporation of a wide range of categories, e.g. animals, foods, and body parts. Even whole events can be incorporated into nouns as verb roots. In this case as well, inter-language comparison suggests that these incorporations developed not at once but successively, newer incorporations following older ones. They also range from productive to idiomatic.

In well-known languages such as English or German, the relatively large variety of modal and semi-modal verbs may stand for the same phenomenon, although surely not with the same abundance. Historically speaking, there was a wave of emergence of the modals from Old to Middle English, and another wave of emergence of the semi-modals from Middle to Modern English. The prior emergence of one or two auxiliary verbs seems to have drawn others along the same path, and the modals as extant structures probably induced the later development of the semi-modals. Krug (2000) explained the development of the semi-modals in terms of a 'gravitational model', which operates on the principle that 'larger masses (in our case highly frequent emerging auxiliaries) attract smaller masses (in our case less common constructions)' (p. 226). Based on frequency and similarity, Krug (p. 232) calculated to which extent specific semi-modals are influencing others. More generally, then, grammatical items and constructions with low frequency change in analogy to high-frequency items and constructions. Type frequency determines the influence of a group of items on other items outside that group. This had already been observed by Paul (1891: 100): 'all that part of language which lacks the support of an environing group, or which enjoys it only in a limited measure, proves, unless impressed by repeated usage intensely upon the memory, not strong enough to withstand the power of the larger groups'.[1] Bybee and Thompson (1997: 384) likewise suggest that

[1] 'Alles dasjenige aber, was die Stütze durch eine Gruppe entbehrt oder nur in geringem Masse geniesst, ist, wenn es nicht durch häufige Wiederholung besonders intensiv dem Gedächtnisse eingeprägt wird, nicht widerstandsfähig genug gegen die Macht der grösseren Gruppen' (Paul 1880: 71).

'high type frequency ensures that a construction will be used frequently, which will strengthen its representational schema, making it more accessible for further use, possibly with new items'. In contrast to type frequency, token frequency may determine the item or construction with the biggest influence within a specific group. In this manner, the dominant structural and functional types of a language exert influence on the paths of new grammaticalizations.

Many chapters in this volume reveal categories that seem to be grammaticalized with preference in certain language groups. A good example may be the stunning range of aspectual and actional categories that is grammaticalized across Turkish languages (Johanson and Csató, Chapter 8 this volume), and the instrumental affixes in many North American languages. Kutenai (Mithun, Chapter 15), for example, has several hundreds of such affixes. Similarly, as Klamer (Chapter 12) shows, Papuan languages have grammaticalized a wealth of applicative prefixes. If one can on the other hand speak of a 'florescence' or preponderance of certain structures as sources, there is no dearth of examples either. For instance, converbs are prominently involved in grammaticalization in Turkish (Johanson and Csató, Chapter 8) as well as other Northeast Asian Transeurasian languages (Narrog, Rhee, and Whitman, Chapter 9); body-part nouns are prominent source structures in South American (Zariquiey, Chapter 17) and Caucasian languages (Arkadiev and Maisak, Chapter 7), while serial verbs serve to grammaticalize a wide range of categories in Timor-Alor-Pantar (Klamer, Chapter 12) and Oceanic languages (Moyse-Faurie, Chapter 14).

However, forces influencing grammaticalizations are not restricted to extant language- or language-area-specific structures. DeLancey (2001) claims that there are functions that are 'important enough, cross-linguistically, that in language which does not formally express it with dedicated grammatical machinery, any construction or lexical means which expresses a related function is a likely candidate for grammaticalization' (p. 15). He labels this kind of cross-linguistically salient function as a 'functional sink'. For example, the formation of adjectives is based on the function of noun modification, which is universal, even if adjectives as a part of speech are not. Certain nouns or verbs may be drawn into this functional sink, eventually leading to the development of a new category adjective in a given language. In the domain of grammar, Thornes (2013) claims that such a functional sink is at work in the grammaticalization of causative constructions in Northern Paiute. In his view, causative is a grammatical function with a high communicative need that is usually available in some form in any language, even in the absence of grammatical means. Because of the frequent need of expression, it attracts lexical materials to grammaticalize.

These sorts of universal aspects of grammaticalization also resonate well with the chapters on creoles in this volume. Both McWhorter (Chapter 19) and Smith (Chapter 18) argue against creole exceptionalism. McWhorter suggests that, 'in terms of the grammaticalization as a process, creoles offer no insights that could not be gained from other languages. No development in Saramaccan (or in other creoles that I am aware of) exemplifies a process, trend, or directionality counter to the grammaticalization process as documented in languages around the world [. . .].' Smith (Chapter 18 this volume) applies accountable quantitative methods and shows that grammaticalization theory can account for patterns we observe in the domain of tense-aspect

expression of Palenquero: 'Palenquero is behaving no differently in the realm of tense and aspect than any other world language, despite its classification as a creole.'

McWhorter (Chapter 19), on the other hand, points to a quantitative rather than qualitative difference: 'Grammaticalization has indeed occurred to an unusually vast degree in the few centuries that most creoles are known to have existed, such that it is reasonable to state that rampant grammaticalization is a defining feature of languages born from pidgins and reconstituted as new languages.' The presumptive reason is that creoles are originally the product of adult language acquisition, and rapid grammatical elaboration took place in later generations. As McWhorter puts it, 'circumstances were ripe for the emergence of entire new paradigmatic systems and overt markings of semantic categories many languages leave to context.'

1.3 GRAMMATICALIZATION AS A POSSIBLE EXPLANATION FOR TYPOLOGICAL FEATURES OF LANGUAGES

As already indicated in the previous section, grammaticalization can also be taken as at least a partial explanation for certain typological features of languages. Three kinds of scenarios can be identified in the literature. First, grammaticalization has been proposed as a partial explanation for some implicational universals; second, grammaticalization has been proffered as a (partial) explanation for the order of affixed material; and third, grammaticalization has been given as an explanation for cross-linguistic types of the expression of certain grammatical categories. Additionally, a number of authors in this volume point to the role of grammaticalization in the creation of analytic (vs synthetic) language structures.

Concerning the first case, Lehmann (1986: 12) observed that the process of grammaticalization can be taken as a causal factor of some for Greenberg's implicational universals of word order, such as (7) and (8).

(7)   In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes.

(8)   Languages with dominant VSO order are always prepositional.

These generalizations can be explained by the fact that when lexical heads of complex constructions grammaticalize (e.g. relational noun to adposition, verb to auxiliary), they usually remain in their original position. Relevantly for the generalizations given above, relational nouns may grammaticalize to adpositions and remain in their original position. However, Lehmann (1986: 13) does not consider grammaticalization as the ultimate cause or motivation for the order and relationship between elements of the sentence, but instead as a channel of change from the lexical to the grammatical category. Thus, grammaticalization does not serve here as a full explanation for the change.

Relatedly, Greenberg himself (1995) referred to grammaticalization within a four-part approach to diachronic typology, which consisted of a dynamicized

state-process model, an elaborate sub-typology, intragenetic comparison, and inter-genetic comparison. Within this approach, Greenberg (pp. 157, 159–60) primarily saw a significant role for grammaticalization in intragenetic comparison, where grammaticalization theory provides knowledge about the directionality of change, but also in intergenetic comparison, where grammaticalization theory does the same on a larger scale. Greenberg was mainly interested in how grammaticalization interacts with global constituent order and word order changes to produce detailed variations in word order within one language, and counterexamples to implicational universals. In the reconstructed stages of word order development in Ethiopian Semitic languages, as represented in (9), the part of the chain shown in bold type constitutes a violation of implicational universal in (7).

(9)   Pr/NG/NA → Pr/NG/AN → **Pr/GN**/AN → Pp/GN/AN (Greenberg 1995: 155;
      A=adjective, G=genitive; N=noun, Pr=preposition, Pp=postposition)

This violation can be explained through the interplay of principles of grammaticalization and global constituent order and word order change in this language. Noun–adjective order is the type of word order that is the least stable and the most susceptible to change. It will change first, followed by noun-genitive order, while grammaticalization from relational noun to adposition will take the most time, and therefore adpositions will lag behind, leading to the apparent violation of universals.

A second case in which grammaticalization is at least partially responsible for typological structures is the order of affixed material. There are at least two related topics in linguistic typology, the suffixing preference, and morpheme order in complex words. It has long been kown that there is an overall tendency in languages to prefer suffixes over prefixes, the 'suffixing preference'. As Bybee et al. (1990) showed, this tendency even holds in the majority of head-initial, especially SVO, languages. Table 1.1 shows the overall preference for suffixation, irrespective of basic word order, according to cross-linguistic data by Dryer (2005).

Although there is also more recent literature, perhaps the most detailed study on the extent of the suffixation preference and its causes is still found in Bybee et al. (1990). In this study, the authors give an overview of the suffixation preference, overall and by

TABLE 1.1. **Prefixing versus suffixing in inflectional morphology** (Dryer 2005: 110)

	No. of languages
Little or no inflectional morphology	122
Predominantly suffixing	382
Moderate preference for suffixing	114
Approximately equal amounts of suffixing and prefixing	130
Moderate preference for prefixing	92
Predominantly prefixing	54
Total	894

word-order type, investigate a number of psycholinguistic (processing) and phonological factors possibly leading to the suffixing preference, and find that none of these factors can account for it. They conclude that the 'fossilized syntax hypothesis' explains the suffixing preference best (Bybee et al. 1990: 35). The 'fossilized syntax hypothesis' goes back to the idea by Givón (1971) that 'yesterday's syntax is today's morphology'. It says that 'that the position of an affix is the same as the position of the non-bound lexical or grammatical material from which the affix developed' (Bybee, Pagliuca, and Perkins 1990: 3). The idea of fossilized syntax led to the revival of interest in grammaticalization from the 1980s.

Strikingly, fossilized syntax does not necessarily reflect normal word order. The most conspicuous example is that person endings are suffixed even in SOV languages, where subjects precede verbs. In response to this problem, it has been hypothesized that it is unstressed pronominal subject pronouns postposed to the verb, rather than pronouns in their normal position, that get grammaticalized (cf. Bybee et al. 1990: 10). Further research has shown that within person paradigms, there is a prefixing preference for very small and very large paradigms, while medium-size paradigms are predominantly suffixing (Cysouw 2009). Cysouw concludes: 'The big riddle of the suffixation preference thus actually consists of various smaller-scale riddles concerning different kinds of affixation asymmetry' (p. 14). Mithun (2003) similarly argues that the ultimate answer to the suffixing preference must be sought in the history (i.e. grammaticalization) of the individual morphemes that as an aggregate make up the suffixing preference.

Generally speaking, we may safely assume that the position of a bound morpheme with respect to the lexical stem reflects the order of the erstwhile independent word vis-à-vis this host, unless one can make the well-founded assumption that a morpheme changed its position after grammaticalization. While Harris and Campbell (1995: 64–70) refer rather abstractly to cases where reanalysis in grammaticalization-like changes led to a change in position, these cases should be considered as exceptional, since bound morphemes are far less mobile than independent ones. However, a change of position of a bound morpheme becomes more likely if the morpheme goes through a clitic stage. As Comrie (1980: 85–7; 1989: 217–18) observed, the position of clitics is freer than that of other bound morphemes, and that for prosodic reasons, clitics may follow rules differing from those for independent words. Thus, if a clitic stage is involved in the grammaticalization of a morpheme, the likelihood rises that morpheme order does not reflect erstwhile word order.

In any case, it does seem that grammaticalization is the most important mechanism behind the suffixing preference. Of course, grammaticalization as such usually cannot explain why an element is in a position before or after a lexical stem at the time when it grammaticalizes.

Beyond the phenomenon of the suffixing preference, morpheme order among affixes in morphologically complex languages is also an intriguing problem that is not fully resolved. But grammaticalization appears to be at least one important motivation, as has been shown by Mithun (2011) in her study on the Navajo verb. The order of morphemes posits a number of difficulties for explanation. For example, (i) languages with verb-final syntactic structure are expected to be suffixing.

Verb-final Navajo, by contrast, is exclusively prefixing. (ii) Mutually dependent morphemes should be contiguous, but in Navajo, some are scattered throughout the verb. (iii) Inflectional affixes are expected to occur further away from the root than derivational affixes, but in Navajo, derivational and inflectional prefixes are interwoven. (iv) Paradigmatically related affixes are expect to occur in the same position in a template. But they do not do so in Navajo. According to Mithun (2011: 179–80), these phenomena are especially challenging. Furthermore, previous attempts at explanation—for example in terms of syntax (the 'mirror principle', Baker 1988), semantic scope (Rice 2000), and a combination of syntactic and phonological principles (Hale 2001)—have failed. Instead, there is good evidence that the order of morphemes can be explained as the order in which these morphemes grammaticalized. Mithun (2011) notes an increase in (1) phonological reduction, (2) generality and abstraction, and (3) diffuse meaning, from left to right in the template. In other words, there is an increasing degree of grammaticalization from left to right. Therefore, 'the positions of prefixes in the verb correlate with their age: those closest to the stem are the oldest, and those furthest the youngest' (Mithun 2011: 184). Hence it is grammaticalization that best explains morpheme order. Thus, in contrast to the case of the suffixing preference, in the case of morpheme order—at least in Navajo—grammaticalization appears to be the immediate cause.

The third area in which grammaticalization has been found to motivate typological patterns has grammaticalization as the source, and hence also the explanation, for cross-linguistically recurring types of expression of certain grammatical categories. Many linguistic categories are cross-linguistically expressed by a limited number of structural types. These structural types in turn are the product of grammaticalization. Furthermore, the source and the degree of grammaticalization of these structures can explain at least some of their morphosyntactic and semantic features. Especially prominent research linking typological patterns with grammaticalization is associated with Heine (1997a; Heine and Kuteva 2002) and Bybee (2006a; Bybee, Perkins, and Pagliuca 1994). In the following, we present a number of examples.

Indefinite articles. According to Heine (2004; Heine and Kuteva 2006), about 89 per cent of all indefinite articles cross-linguistically are derived from the numeral 'one'. This explains some positional tendencies of indefinite articles, the fact that they are often confined to singulars, and the following implicational hierarchy for their application: mass noun > plural noun > singular noun.

Possessive constructions. Heine (1997a) identified the eight cross-linguistic source schemas for possessive constructions that are listed in (10).

(10) 1. Action   *X* takes *Y*
      2. Location   *Y* is located at *X*
      3. Companion   *X* is with *Y*
      4. Genitive   *X*'s *Y* exists
      5. Goal   *Y* exists for/to *X*
      6. Source   *Y* exists from *X*
      7. Topic   As for *X*, *Y* exists
      8. Equation   *Y* is *X*'s *(Y)*

Crucially, these source schemas can account for some characteristics of specific possessive constructions, such as why they often have non-verbal, or copular-like, predicates, how the possessor is encoded in a specific language, i.e. as a comitative, locative, etc., or why they frequently have locative morphology etc. (cf. Heine 1997a: 102–3). Furthermore, these constructions often undergo a development at the end of which (1) the possessor precedes the possessee, (2) the possessor has properties of a subject, and the possessee has properties of a clausal object, (3) the possessor is definite and the possessee is indefinite. Therefore, possessive constructions often display 'hybrid' properties between source and target structures depending on the stage of development (Heine 1997a: 98–9). In this manner, grammaticalization as a process can account for the properties of these constructions.

Future. In her analysis of the cross-linguistic polysemy of future morphemes, Bybee (1988) found that the polysemy can be explained by reference to their diachronic evolution, which takes place in the form of paths, such as from movement ('go' or 'come') to intention, then to prediction, the core future meaning, and further to other meanings such as supposition or imperative. According to Bybee (1988: 374–5), such paths of development are explanatory because they explain (1) why it is difficult to find a single abstract meaning for a polysemous morpheme (like many future morphemes), (2) the cross-linguistic similarities of grammatical meanings by similar paths of development and principles of historical change, and (3) differences between morphemes in different languages with reference to different lexical sources and differing extent of change along the universal paths of change. Furthermore, (4) they make it possible to predict possible combinations of meanings, and (5) they allow the reconstruction of the lexical sources of grammatical morphemes.

Passives. According to Givón (2008, 2009), six common types of passive constructions can be identified: (1) the adjectival-stative passive (e.g. English common passive); (2) the reflexive passive (e.g. English *get*-passive); (3) the serial-verb adverse passive (e.g. Chinese); (4) the VP-nominalization passive (e.g. Ute); (5) the left-dislocation-cum-impersonal-passive (e.g. Kimbundu); (6) the zero-anaphora passive (e.g. Sherpa). Types (1)–(3) are so-called promotional passives (i.e. the erstwhile object is promoted to subject), while (4)–(6) are non-promotional. Note that the concept of passives applied here is fairly broad; not every study of passives would include the same range of constructions. In any case, the concrete structural properties of these passive structures in individual languages can be explained by the degree to which they have grammaticalized to more prototypical passives. For example, non-promotional passives can eventually become promotional as subject properties gradually shift to the object-patient. Similarly, oblique agents may eventually be added. In this way, types of passives and their morphosyntactic features can be explained with reference to their source construction and their degree of grammaticalization.

Heine (Chapter 2 this volume) refers to a number of avenues of grammaticalization that are prevalent in Africa and belong to limited sets of schemas accounting for the large majority of grammaticalizations of specific grammatical categories cross-linguistically. One example of this process is verbs of action being appropriated for the expression of comparison; another is de-andative futures.

Besides the cross-linguistically common schemas, some chapters also discuss typical sources for various grammaticalizations such as posture verbs for aspectual meanings and 'say'-verbs as quotatives (Arkadiev and Maisak, Chapter 7 this volume; Moyse-Faurie, Chapter 14), adversative constructions for passives (Coupe, Chapter 10), or verbs of transfer marking beneficiaries or recipients (Chapter 14). But there is also the occasional outlier like the polygrammaticalization of 'return' into reflexive and reciprocal markers, prepositions ('until') and conjunctions ('then'), or the grammaticalization of 'go down' into reflexives and reciprocals in Oceanic languages (Chapter 14).

A topic brought up in several contributions to this volume is the role of grammaticalization in the change from synthetic to analytic structures, and back to synthetic structures over longer periods of time, and relatedly, the genesis of inflections as typical synthetic structures. Haspelmath (Chapter 6), alluding to historical linguists of the 19th and early 20th century, coins the term 'anasynthetic spiral' for this large-scale type of change. While its application to languages as a whole is controversial, the occurrence of anasynthetic change in specific categories in specific languages is much easier to demonstrate. Haspelmath brings up a number of examples and possible motivations for the anasynthetic spiral, and concludes that the extravagance and inflation model of grammaticalization is best suited to account for it. Likewise, Esseesy (Chapter 3) states that '[g]rammaticalization has been shown to facilitate the change from the direction of synthetic to analytic in several Semitic languages [. . .] and facilitates the transition from one state to another and in some cases perhaps back in a cyclical fashion'. Haig (Chapter 4) also emphasizes the role of grammaticalization in cyclical typological change in Iranian languages, suggesting that 'the history of grammaticalization can to some extent be seen as the gradual re-acquisition of lost morphological categories'.

Part of Haspelmath's study is the presumptive role of a flectional-fusional stage leading from agglutinating to isolating structures, and he bemoans the fact that cross-linguistically the origins of flectional/fusional patterns are mostly unknown. This topic is echoed in other chapters in a more concrete form. Haig (Chapter 4), looking at historical data from Iranian languages, concludes that 'inflectionalization is evidently a process that requires millennia, not centuries, to achieve'. That is, historically observing an entire process of inflectionalization from lexical item to inflectional ending would require a time-depth of data that extant historical records of languages do not afford us. Thus, Haig writes, 'the assumed final stage of grammaticalization, namely into fully-fledged inflection, is an exceedingly slow process indeed, taking millennia before all traces of the lexical, or at least non-inflectional, origins of grammatical formatives are lost.' This may be one reason why even in Indo-European languages, which have been the cornerstone of grammaticalization research, very few inflectionalizations have been observed historically, as Dahl (Chapter 5) remarks, thus relativizing the contrast between East mainland and Southeast Asian languages with little morphological grammaticalization, on the one hand, and European languages, on the other.

Likewise, pointing to the required time depth for morphological grammaticalizations, Mushin (Chapter 13), points out with respect to Australian languages: 'Few

grammatical categories are regularly marked by forms whose lexical source is still available as a free form. [...] It is therefore challenging to find clear comparative evidence of contemporary bound affixal forms that in some languages may retain features of their lexical origins.' Often, the best that comparative evidence can provide is to identify different stages in the process of grammaticalization of the same categories (in this case, clitic constructions) in related languages.

The slow pace of the genesis of inflections contrasts with a potentially rapid pace of their decay. With respect to the Iranian languages, Haig (Chapter 4) concludes: 'Inflectionalization is evidently a process that requires millennia, not centuries, to achieve, though paradoxically, its loss can be quite rapid, even catastrophic.'

While morphologization and especially inflectionalization may generally take a very long time, the emergence of the expression of new grammatical categories through independent morphemes or clitics can be rather quick given the right circumstances, as shown by the high pace of grammaticalization in creole languages (Smith, Chapter 18, and McWhorter, Chapter 19).

Beyond the specific issue of creoles, it is clear that the rapid emergence of new categories, or a profound change in the typological profile of a language, is more likely to be brought about by intense language contact than by the primarily language-internal changes that are at the core of the idea of an anasynthetic spiral. Here as well, grammaticalization plays an important role. Cases in point are the development of classifiers (through grammaticalization) in Papuan languages (Klamer, Chapter 12), and of evidentials as a category in Tariana (Aikhenvald, Chapter 16), an Arawak language in strong contact with Tucanoan languages.

1.4 CONCLUSION AND ORGANIZATION OF THE BOOK

The studies in this book deal both with cases where typological features of language apparently influence grammaticalization paths and with cases in which grammaticalization creates typological features. As for the first, variation in grammaticalization is most obvious with respect to phonological and morphological aspects of grammaticalization, but may also pertain to syntax and semantics. The most striking case where typological features of a language constrain morphological and phonological grammaticalization is still the tendency of isolating languages not to develop affixal material and grammatical paradigms (Ansaldo, Bisang, and Szeto, Chapter 11). However, the extent to which grammaticalization differs in these languages may not be as great as is sometimes thought (cf. Dahl, Chapter 5). Furthermore, if the core of grammaticalization is semantic (functional) change, as argued in Heine (Chapter 2) and Narrog (2017b), then the morphological aspects are more peripheral, although still of interest. On the semantic and syntactic side, it seems that there is generally a tendency in languages to follow already trodden grammaticalization paths and reproduce or flesh out established grammatical categories, often to a considerable extent, rather than to create entirely new structures and categories. A salient departure from this tendency towards conservatism, or inertia, is most likely to take place under intense language

contact, as has been demonstrated perhaps most clearly in Akhenvald's (2002) analysis of Amazonian languages.

As for the case of grammaticalization creating language structures of typological relevance, it has been found that the order of grammatical morphemes can be explained by their position at the time of their grammaticalization. This seems to hold for morphologically bound as well as for non-bound morphemes. Second, commonalities and divergences in the coding of grammatical categories across languages seem to be motivated to a large degree by grammaticalization. Thus, as Bybee (2009: 30) put it, 'grammaticalization has great potential for explaining the similarities as well as the differences among languages'. A third area where grammaticalization strongly contributes to typological features of languages is the cycle—or spiral—between analytic and synthetic language structures that clearly takes place in the life of specific categories in specific languages, but perhaps even at the level of overall structure of a specific language. This cycle/spiral feeds on the mechanisms of grammaticalization. A complicating factor in documenting full cycles or spirals is that while the decay or disappearance of extant morphological marking can be very quick, its development may take a very long time, especially when it comes to full-fledged morphologization such as inflectionalization. Lastly, grammaticalization is an important player when languages develop new categories or a new typological profile under intense language contact. There is a strong tendency for languages in contact to adopt features from each other and thus develop a similar typological profile. This is a major reason why we decided to organize this book mainly in terms of language areas.

The order of chapters in this book starts with Africa and then proceeds towards the north and the east, roughly but not exactly mimicking the possible spread of mankind. The chapters on African and Iranian languages by Heine, Esseesy, and Haig are followed by chapters on European and Caucasian languages. The chapter by Haspelmath is classified as a 'European' chapter although it tackles a general issue, because this general issue has grown out of traditional European linguistics and evidence has been brought up mainly from European languages. We then proceed towards Turkic languages (Johanson and Csató, Chapter 8), and as an eastward extension, part of the group of so-called Transeurasian languages, Korean and Japanese (Narrog, Rhee, and Whitman,Chapter 9). Moving a step back westwards, Coupe (Chapter 10) deals with South Asian languages, and Ansaldo, Bisang, and Szeto (Chapter 11) with issues in East Asian mainland and Southeast Asian languages. These contributions are followed by a cluster of chapters on Papuan (Klamer, Chapter 12), Australian (Mushin, Chapter 13), and Oceanic (Moyse-Faurie, Chapter 14) languages. Finally, across the Pacific, three chapters deal with indigenous American languages—Mithun (Chapter 15) for North America, and Aikhenvald (Chapter 16) and Zariquiey (Chapter 17) for South America—and the last two chapters by Smith and McWhorter have (American) creole languages as their topic.

2

Grammaticalization in Africa
Two contrasting hypotheses

BERND HEINE

2.1 INTRODUCTION

2.1.1 A NOTE ON METHODOLOGY

Work on grammaticalization is based on historical reconstruction, and the safest way to achieve reconstruction is by drawing on historical documents that provide information on earlier states of language use. However, restricting the study of grammaticalization to written languages would mean that more than 95 per cent of the world's languages would have to be excluded. We therefore adopt also an alternative but well-established methodology of reconstruction that has been employed mostly for unwritten but also for written languages. This methodology relies mainly on three components: (a) diachronic reconstruction, e.g. by means of the comparative method, (b) internal reconstruction, and (c) typological generalization.

The following example may illustrate this methodology (see also Heine 2003: 580). The Bantu language Swahili of eastern Africa has a future tense prefix *-ta-*, which is hypothesized to be historically derived from the volition verb *-taka* 'want' on the basis of the following evidence. By using (a) it is possible to establish that the verb must be older than the future tense marker: The application of the comparative method shows that the verb *-taka* is a modern reflex of the Proto-Bantu verb **-càk-a* 'desire', while it is not possible to reconstruct the future tense marker back to Proto-Bantu (Guthrie 1967–71). (b) Internal reconstruction suggests, for example, that the earlier form of the tense marker is likely to have been *-taka-* since the form *-taka-* is still retained in relative clauses. Method (c) allows for two kinds of generalization. First, it establishes that verbs of volition ('want', 'desire') quite commonly give rise to future tense markers in the languages of the world, the English *will*-future being a case in point (see Heine and Kuteva 2002, WANT > FUTURE). And second, processes of this kind tend to involve specific types of semantic, morphosyntactic, and phonological change: loss of lexical in favour of grammatical meaning (desemanticization),

loss of morphosyntactic properties, such as loss of word status (decategorialization), and loss of phonetic substance (erosion).

On the basis of these methodological tools it is possible to formulate a strong hypothesis to the effect that the Swahili future tense marker -*ta*- is the result of a common grammaticalization process, having lost its lexical meaning of volition (desemanticization), its status as an independent verb (decategorialization), and part of its phonetic substance, being reduced from -*taka* to -*ta*- (erosion). To conclude, while it is always desirable to search for historical records, we argue that such records are not a requirement for the reconstruction of grammaticalization processes.

2.1.2 GRAMMATICALIZATION STUDIES ON AFRICAN LANGUAGES: AN OVERVIEW

For over 30 years now, African languages have been the subject of studies in grammaticalization (e.g. Heine and Reh 1984; Heine and Claudi 1986; Heine and Hünnemeyer 1988; Heine et al. 1991a, 1991b; Heine 1994a, 1997a, 1997b, 1997c, 2000, 2011a; Heine and König 2005; Heine and Miyashita 2008; Heine and Narrog 2010). As all these studies suggest, many of the pathways of grammaticalization that have been recorded from other parts of the world are also documented in Africa (see Heine and Kuteva 2002).

These studies were based mostly on a comparative methodology. Much of the work aimed at contributing to the description of African languages also had a typological perspective. Underlying this work there were a number of goals, but clearly the main goal was to explain language structure and to search for typological regularities. Since language structure is a product of language use in the past, explanations were sought mainly in the diachronic development of grammar (Heine 1997a).

This work resulted in a number of different kinds of publication. First, there is a monographic treatment of grammaticalization in African languages in Heine and Reh (1984). Second, there are a number of general typological studies that include but are not restricted to grammaticalization in African languages (Heine et al. 1991a; Heine 1993, 1994a, 1997a, 1997b; Heine and Kuteva 2002, 2005, 2006, 2007). And third, there are typological studies each dealing with a specific domain of grammar or grammatical function. Specifically, the domains and functions focused on were: from compounding to derivation (Heine and Hünnemeyer 1988; Heine et al. 1991a, 2015; Heine and Kuteva 2009); from noun to adposition (Heine et al. 1991b), from verb to auxiliary (Heine 1993); from verb to complementizer (Lord 1976); the grammaticalization of serial verbs (Hünnemeyer 1985; Lord 1993); reflexives and reciprocals (Heine 2000; Schladt 2000; Heine and Miyashita 2008); comparative constructions of inequality (Heine 1997a; Leyew and Heine 2003); verbal proximative aspects (Heine 1992, 1994a, 1997c); the metaphorical basis of grammmaticalization (Claudi and Heine 1986; Heine et al. 1991a; Mkhatshwa 1991); and grammaticalization chains as linguistic categories (Heine 1992).

Finally, a considerable part of this work was dedicated to understanding the role that language contact has played in shaping the areal landscape of the African

continent. This work aimed, on the one hand, at defining Africa as a whole as a linguistic area (Heine and Leyew 2008; Güldemann 2005; Clements and Rialland 2008; Creissels et al. 2008). On the other hand, it was driven by a search for areal patternings within Africa cutting across genetic (genealogical) boundaries (Heine 1994a; Kuteva 2000; Leyew and Heine 2003; Güldemann 2005, 2008a; Kießling, Mous, and Nurse 2008; Heine 2011b; Heine and Kuteva 2011).

   Despite the fact that there are hardly any historical documents on earlier states of African languages, research carried out in the course of previous decades demonstrates not only that it is possible to reconstruct grammatical evolution but also that grammaticalization theory can be of help in defining processes leading to areal diffusion in Africa. With regard to our understanding of areal diffusion, findings made in this work can be summarized thus. First, it is possible, at least on a quantitative basis, to distinguish the languages of Africa from those in other parts of the world (Heine and Leyew 2008). Second, there are a few linguistic macro-areas in Africa, most of all the Ethiopian area (for a summary, see Crass and Mayer 2008, 2009) and the Macro-Sudanic Belt (Güldemann 2008a). Second, there are also some micro-areas, such as the Tanzanian Rift Valley (Kießling et al. 2008). And third, in all this work on areal patternings in Africa, findings on grammaticalization have played an important role (see especially the contributions in Heine and Nurse 2008).

2.1.3 TWO HYPOTHESES

As the work alluded to in the preceding section suggests, grammaticalization is a well-researched topic in African linguistics—more than in some other regions of the world. The present chapter will deal with grammaticalization processes in general and more specifically with the relationship between form and meaning in such processes. To this end, we will look at African language data in order to evaluate two hypotheses that have been proposed on this issue. These hypotheses, which we will refer to as the parallel reduction (PR) hypothesis and the meaning-first (MF) hypothesis, are now looked at in more detail.

2.1.3.1 The parallel reduction hypothesis

According to a widespread assumption, going back to the early phase of modern grammaticalization studies (Bybee and Dahl 1989; Bybee, Perkins, and Pagliuca 1994: 20, 107; Lehmann 1982: 123), meaning and form proceed in parallel—that is, there is coevolution, captured appropriately by the parallel reduction hypothesis of Bybee et al. (1994: 107). This hypothesis, henceforth called the PR hypothesis, can be summarized in short as in (1).[1]

---

[1] The term 'grammaticalization' has been employed for a wide range of linguistic changes. In the present chapter we are restricted to 'paradigm' cases that arise via the the evolution of grammatical categories expressing schematic functions relating to tense, aspect, modality, evidentiality, number, gender, (in) definiteness, case, subordination, etc. The term 'meaning' is used here in contrast with 'form', i.e. 'meaning' in this sense also includes pragmatically induced factors. That it is useful to distinguish semantics from

(1) The parallel reduction hypothesis
     Form change parallels meaning change in grammaticalization

Bybee and associates found that 'form and meaning covary in grammaticization on a large body of data' (Bybee et al. 1994: 20; see also Bybee, Pagliuca, and Perkins 1991). Distinguishing two types of form change, namely the reduction or loss of phonetic bulk and the fusion of the grammaticalizing material to surrounding material, the authors found that 'both types of formal change in grammaticization parallel the main types of semantic change in grammaticization' (Bybee et al. 1994: 106). On the basis of substantial cross-linguistic data on the evolution of tense, aspect, and modality, Bybee et al. (1994: 106–15) in fact provide strong evidence in favour of the hypothesis in (1). This evidence is based on typological analysis and comparison of data from 76 languages in a carefully chosen sample of the languages of the world. The data rest on grammatical descriptions of established grammatical forms in the languages concerned—i.e. on more or less *conventionalized* grams—or, in other words, on established grammatical categories.

This coevolution hypothesis has provided an important generalization on the evolution of grammaticalization. But its scope in explaining grammaticalization is restricted, as some lines of research suggest, most of all that by Bisang (2004; see also Bisang 2011). Bisang concludes that East Asian and mainland Southeast Asian languages represent a type of grammaticalization that is characterized by its limited coevolution of meaning and form (Bisang 2011: 116; but see also Ansaldo and Lim 2004). Note also that, as observed by Narrog (2017b: 105–6), there is evidence to the effect that form change and function change in grammaticalization do not share the same motivation, and that 'formal grammaticalization as such cannot be regarded as essential for grammaticalization'.

2.1.3.2 The meaning-first hypothesis

The PR hypothesis contrasts with that proposed by Heine, Claudi, and Hünnemeyer (1991a: 28; see also Heine 1997a: 3–4) according to which change in meaning precedes change in form in grammaticalization; let us refer to this as the MF (meaning-first) hypothesis.[2]

Observations in support of this hypothesis can be found already in some of the work on grammaticalization in the 1970s and 1980s (e.g. Givón 1971, 1975; Lord 1973), and such observations have also been made in some form or other in more recent work; cf. the extravagance hypothesis of Haspelmath (1999) or the extra-clarity hypothesis of Michaelis and Haspelmath (2015). The hypothesis is suggested most of all by work written in the tradition of Heine et al. (1991a; see also Heine

---

pragmatics has been documented abundantly in the relevant literature, including the literature on grammaticalization (e.g. Bisang 2011).

  [2] This hypothesis covers also cases of grammaticalization that have undergone changes in meaning but not in form (Bisang 2004; see below), and where it would be correct to say 'meaning only' rather than 'meaning first'. I am grateful to Heiko Narrog (p.c.) for having drawn my attention to this observation.

1997a), which is firmly based on the MF hypothesis. In this tradition it is argued that the main motivation underlying grammaticalization is to communicate successfully. One salient human strategy consists in using linguistic forms for meanings that are concrete, easily accessible, and/or clearly delineated to also express less concrete, less easily accessible and/or less clearly delineated meaning contents. To this end, lexical or less grammaticalized linguistic expressions are pressed into service for the expression of more grammaticalized functions (Heine, Claudi, and Hünnemeyer 1991a: 28; Heine 1997a: 4–7, 2003; Narrog and Heine 2017).[3] The only reasonable conclusion to be drawn from this hypothesis is that interlocutors are first concerned with what they say, i.e. with meaning, before changing their habits on how they say what they say—hence, semantic change is assumed to precede formal change. On this position, the MF hypothesis can be formulated as in (2):

(2)   The meaning-first hypothesis
      Semantic change is primary in grammaticalization and precedes form change
      (i.e. morphosyntactic and phonological change) in time

That (2) is correct has also been argued in Heine, Kuteva, and Narrog (2017). According to the observations made there, the only unambiguous factor that appears to account for directionality in grammatical change is the semantic relation between the source structure and the target structure, where the former is frequently but not necessarily a lexical structure. As this study suggests, other than the semantic relation between source and target there does not seem to be any other factor, such as contextual features, inferential mechanisms, analogy, or constructional form, that ultimately can be held responsible for unidirectionality in the history of, e.g., the English *be going to* future.

   Hypothesis(2) has also implicitly or explicitly been claimed in a number of other studies, where it is argued that semantic change drives form change in grammaticalization (e.g. Fischer 2000: 155; see Börjars and Vincent 2011: 168). It forms a central assumption of the framework of Heine et al. (1991a: 28–9) and Heine (1997a: 4–7), where grammaticalization is viewed essentially as a cognitive-communicative and semantic process. Accordingly, explaining this process must be first of all with reference to meaning. That it is the meaning (or function) that drives grammaticalization has also been suggested in studies concerned with discourse functions, such as Harder and Boye (2011: 65), who identify as a prerequisite for grammaticalization the relative usefulness of source expressions for 'a discursively secondary role'. This usefulness accounts for their frequency and subsequent conventionalization in their secondary role.

---

[3] This hypothesis is outlined in Heine et al. (1991a): 'there is one specific principle that can be held responsible for the creation of linguistic forms serving the expession of grammatical concepts.' This principle is referred to by Werner and Kaplan (1963: 403) as 'the principle of the exploitation of old means for novel functions'. By means of this principle, concrete concepts are employed in order to understand, explain, or describe less concrete phenomena. In this way, clearly delineated or structured entities and non-physical experience are understood in terms of physical experience, time in terms of space, cause in terms of time, or abstract relations in terms of physical processes or spatial relations (Heine et al. 1991a: 28).

Most commonly, the MF hypothesis is expressed implicitly rather than explicitly. For example, Michaelis and Haspelmath observe:

Grammaticalization involves (i) **semantic change**, which often results in (ii) **functionalization** (content item > function item), and then (iii) **compaction** (cliticization, agglutination, fusion). (Michaelis and Haspelmath 2015: 1; bold print in the original)

This depiction appears to be in accordance with the MF hypothesis in that it implies that semantic change and 'functionalization' precede formal changes in the process.

2.1.4 THE PRESENT CHAPTER

In spite of all the work that has been done on grammaticalization in the course of the last decades, there does not seem to be conclusive evidence to decide which hypothesis is correct. This chapter is restricted to linguistic observations as they can be made in grammaticalization processes commonly observed in African languages.

Grammaticalization is as a rule a long process, extending over decades and centuries, and it is influenced by many factors. Furthermore, the entry point in grammaticalization and the pace of development differ from one marker to another, and from one construction to another (Narrog 2005: 697; 2012: 107–9). Accordingly, when we study the evolution of some grammatical category we are confronted with a long and a complex history, which is not necessarily restricted to language-internal factors but may as well involve language contact (Heine and Kuteva 2003, 2005). This history is to a considerable extent a process of semantic, morphosyntactic, and phonological reduction, as captured by diagnostic techniques such as those of Lehmann (1982: 123), Hopper (1991), and Heine and Kuteva (2002: 2; see also Heine and Narrog 2010: 405). But comparing the semantic and formal structure of modern grammatical categories with that of their non-grammaticalized sources may therefore tell us little about the motivations responsible for the rise of such categories.

The question to be addressed here, therefore, does not concern the overall evolution of grammatical categories—a topic that has aptly been covered in work such as that by Bybee and associates alluded to above; rather, our interest is with the motivations of interlocutors that can ultimately be held responsible for this evolution. We will therefore look at evidence from African languages that allows us to evaluate the two hypotheses. If the PR hypothesis is correct, form and meaning should generally go together—i.e. once grammatical change has taken place it should simultaneously have involved both. By contrast, if the MF hypothesis is correct, there should be evidence to show that there was semantic change but no corresponding form change in grammaticalization. More specifically, there should be language data to show that at the earliest stage of the process it is the former that has taken place while the latter hasn't (yet).[4] Such data can be found if there are

---

[4] We are not aware that an opposite hypothesis to the effect that formal change generally precedes semantic change in grammaticalization has been proposed. Hence, we do not pursue this possibility in this chapter. An anonymous reader, however, draws our attention to a claim first made in the 19th century that analytic formations in Romance languages were a response to the phonetic reduction of Latin grammatical markers.

TABLE 2.1 **The context model of grammaticalization** (where source meaning=non-grammaticalized, temporarily prior; target meaning=new grammatical meaning derived from the source meaning; Heine 2002)

Stage	Context	Resulting meaning
I Initial stage	Unconstrained	Source meaning
II Bridging context	A specific context giving rise to an inference in favour of a new meaning	Target meaning foregrounded
III Switch context	A new context which is incompatible with the source meaning	Source meaning backgrounded
IV Conventionalization	The target meaning no longer needs to be supported by the context that gave rise to it; it may be used in new contexts	Target meaning only

examples of grammaticalization where there was a change from e.g. lexical to grammatical meaning but no accompanying formal change—that is, where there is ambiguity between the two kinds of meaning while their form is (still) the same.

What kind of data these are can be illustrated by means of the context model proposed in Heine (2002), depicted in Table 2.1.

According to this model, the trigger of grammaticalization can be seen in Stage II (and to some extent also Stage III) situations of grammatical change, where in a specific context a linguistic expression is enriched by the rise of a new meaning (the target meaning) and this change does not affect the form of the expression concerned. For the present study, therefore, which is concerned with the reconstruction of the motivations underlying grammaticalization, Stage II and Stage III situations appear to be of paramount importance. The data used by Bybee et al. (1994: 106–15) are for the most part not of this kind; they typically concern conventionalized grammatical categories as they surface in reference grammars—i.e. Stage IV situations. As we will see in section 2.3, this has a bearing on the results obtained.

For the purpose of the following discussions, 'change' will be said to obtain whenever a linguistic expression exhibits a recurrent feature that was absent in an earlier use of the same expression. This feature is semantic in the case of meaning change, while in form change the feature is phonological, morphological, or syntactic, where phonological change includes both segmental and suprasegmental features.

2.2 CASE STUDIES FROM AFRICAN LANGUAGES

The following is a survey of four kinds of grammaticalization commonly found in African languages. The goal of this discussion is to evaluate the hypotheses proposed in section 2.1.3.

2.2.1 DE-VOLITIVE PROXIMATIVES

A grammaticalization process that is cross-linguistically widespread but has been reported to be particularly common in Africa (Heine 1997c) concerns proximative aspect forms and constructions. The function of proximatives is to denote the temporal phase immediately preceding the initial boundary of the situation described by the main verb, common English paraphrases being 'be about to do', 'being on the verge of doing', or 'nearly, almost' (Heine 1992, 1994a, 1997c; Romaine 1999). Presumably the most common though not the only pathway (see Heine 1997c) is one involving the auxiliation of a verb of volition ('want', 'desire', etc.) where this verb turns into a proximative marker while the complement of this verb in the source structure turns into the new main verb in the grammaticalized construction. Example (3), from Swahili of East Africa, illustrates the process concerned: (3a) illustrates the lexical source construction, where the verb -*taka* has the lexical meaning 'want'. This meaning still exists in (3b), but since dying is not something that one normally wants, the grammatical meaning of proximative is foregrounded. Accordingly, in such contexts there is a proximative meaning 'be about to' of the bridging Stage II. This meaning is the only one in contexts where the subject referent is inanimate, i.e. where the semantics of this referent rules out the lexical source meaning of volition, as in (3c). Hence Swahili has also developed a switch Stage III meaning where the proximative provides the only reading and the source meaning is backgrounded, even if it may still be recoverable, for example, in metaphorical interpretations.

(3) Swahili (Bantu, Niger-Congo; Heine 1997c: 5)
 a. Ni-   li-   taka ku-   m-   piga.
    1.SG- PST- want INF- 3.SG.OBJ- hit
    'I wanted to hit him.'

 b. Ni-   li-   taka   ku-   fa.
    1.SG- PST- want   INF- die
    (i) 'I wanted to die.'
    (ii) 'I nearly died. I narrowly escaped death.'

 c. Mvua   i-   li-   taka   ku-   nyesha.
    rain   it-   PST- want   INF-   rain
    'It was about to rain.' (*'The rain wanted to rain.')

In accordance with the context model of Table 2.1, we interpret the three examples in (3) as each representing a different stage of grammatical evolution, where (3a) illustrates the lexical source of Stage I, (3b) the bridging Stage II, and (3c) the switch Stage III. The grammaticalization process illustrated by this Swahili example was restricted to the manipulation of meaning in context; it did not affect the morpho-syntactic or phonological forms, which both remained essentially unaffected. To conclude, already at Stage II there must have been (optional) meaning change not accompanied by change in form—in accordance with the meaning-first (MF) hypothesis in (2), but not with the PR hypothesis in (1). The process reconstructed

above has not proceeded beyond Stage III—i.e. there is no conventionalized Stage IV construction in Swahili.

Examples of proximatives like the one illustrated in (3) are legion in African languages; they can be said to present weakly grammaticalized categories since they have not proceeded beyond Stage III. But there are also examples in African languages where the process has proceeded further, giving rise to fully grammaticalized categories of Stage IV. We may illustrate this with an example from a language not genetically related to Swahili, namely the Maa language of the Nilotic family. The data are taken from the Chamus dialect of north-central Kenya. This dialect appears to have gone through the same stages II and III but has gone one step further, resulting in a full-fledged proximative category. Example (4a) illustrates the Stage I source construction involving the volition verb *-yyéú* 'want' and (4b) the Stage III target meaning of the process, where this construction occurs with inanimate subjects.[5]

(4)   Chamus (Maa dialect, Eastern Nilotic, Nilo-Saharan; Heine 1992: 339)
    a. k-   e-   yyéú   m- partút.'[6]
       k-   3.SG- want   F-   woman
       'He wants a woman/wife.'

    b. k-   é-   yyeu   l-   cáni   n-   éuróri.
       k-   3.SG- want   M-   tree   NAR-   fall
       'The tree almost fell.' (lit. 'The tree wanted to fall.')

Both stages exhibit essentially the same formal structure, once again in support of hypothesis (2): there has been semantic chance but no formal change. But Maa speakers have proceeded beyond these stages: subsequently there has also been formal change, in that the erstwhile verb form *k-e-yyéú* 's/he wants' developed into an invariable proximative aspect particle *(k)éyyeu*, illustrated in (5). Thus, this verb form has undergone internal decategorialization, turning into a frozen particle; in this capacity it is exclusively a proximative marker, which can equally take inanimate and human subject referents. Thus, in addition to retaining the earlier Stage II and Stage III structures, grammaticalization has also led to a fully conventionalized Stage IV construction, where the lexical source meaning of volition is ruled out.

(5)   Chamus (Maa dialect, Eastern Nilotic, Nilo-Saharan; Heine 1992: 339)
    kéyyeu   a-   ók   nánʋ   kʋlέ.
    PROX   1.SG-   drink   1.SG.N   milk.A
    'I was about to drink milk.'

To conclude, Chamus has acquired a new grammatical category, namely an aspect marker, via the grammaticalization of a verb of volition inflected in its third person singular imperfective form, but the new construction, illustrated in (5), coexists with the earlier, weakly grammaticalized construction in (4b). The rise of the Stage IV construction in (5) had dramatic consequences for the morphosyntactic format of the

---

   [5] We are ignoring the tonal inflections to be observed in the following examples, which are morphophonologically conditioned and need not concern us here.
   [6] The prefix *k-* is restricted to the imperfective paradigm of verb forms; its exact meaning is unclear.

construction, which need not concern us here (see Heine 1992: 339 for details). Suffice it to mention the following morphosyntactic change: whereas the inflected verb form *k-e-yyéú* 's/he wants' requires the following main verb to be encoded in the narrative tense (using the narrative inflection *n-*; cf. (4b)), the aspect particle *(k)éyyeu*, illustrated in (5), takes the verb in the unmarked main clause syntax. But what is of interest here is the fact that semantic change must have preceded formal change, which may be taken to suggest that the latter is an epiphenomenal effect of the former.

2.2.2 FROM BODY-PART NOUN TO REFLEXIVE MARKER

The de-volitive proximative categories looked at in section 2.2.1 are in no way exceptional, as can be shown with many other kinds of grammaticalization process. In the present section we look at another process, which concerns the evolution of reflexive markers. A survey of reflexive constructions in the languages of the world suggests that reflexive markers are mainly the product of the grammaticalization of four kinds of conceptual processes, which are based on the strategies listed in Table 2.2.

Our concern here is exclusively with the noun strategy, which seems to be of universal significance but is more widespread in Africa than elsewhere (Heine 2000; Schladt 2000). In accordance with this strategy, noun phrases consisting of a body (-part) noun, usually taking a coreferential possessive modifier, are grammaticalized to reflexive markers when serving as arguments. The source noun is in most cases 'body', less commonly also nouns for 'head', and this situation does not seem to be dramatically different in other parts of the world, as the percentages in Table 2.3 suggest.

TABLE 2.2 **The main strategies to develop reflexive markers (Heine 2000, 2005)**

	Label	Strategy
a	Pronoun strategy [uR] = 'unmarked reflexive'	Use personal pronouns
b	Intensifier strategy	Add an 'intensifier' to (a)
c	Noun strategy	Use a 'body' noun
d	Non-transparent	(Unknown strategy)

TABLE 2.3 **Nominal sources of reflexive markers in Africa and elsewhere (Schladt 2000: 112; Heine, own data from 46 African languages, 49 forms)**

Nominal source	Africa		Other continents		Total
Nominal source	Frequency	%	Frequency	%	
'body'	25	61.0	71	79.8	96
'head'	6	14.6	13	14.6	19
'soul/life'	4	9.8			4
Other body parts	6	14.6	5	5.6	11
Total	41	100	89	100	130

The following example from the Efik language of southeast Nigeria illustrates this pathway: (6a) is characteristic of the lexical source structure of Stage I, where the complement *ídém* 'body' is a noun. Example (6b), by contrast, is suggestive of the switch stage III, where in the context involving the main verb *ńdíwòt* 'kill' the intended meaning of the phrase *ídém ésiĕ* ('her body') is no longer nominal but rather reflexive. In such contexts the lexical Stage I meaning is backgrounded and the grammatical, reflexive meaning foregrounded, while the form is still that of the lexical source construction. Once again we see that semantic change precedes formal change in grammaticalization.

(6) Efik (Benue-Congo, Niger-Congo; Essien 1982)

    a. Árìt   éyĕ   ídém.
       Arit   has   body
       'Arit has a beautiful body.' (lit. 'Arit has body.')

    b. Árìt   óyòm   ńdíwòt   ídém   ésiĕ.
       Arit   want   kill   body   her
       'Arit wants to kill herself.' (lit. 'Arit wants to kill her body.')

But in many African languages the process has advanced one step further, giving rise to a conventionalized Stage IV construction where formal change has also now taken place and the lexical source meaning of Stage I is no longer available. We may illustrate this situation with the following example from Yoruba of southwest Nigeria, which exhibits the whole range of stages of grammaticalization, as the description by Awolaye (1986) suggests.

In example (7(i)), the noun *ara* 'body' (plus its possessive modifier *wọn* 'their') is interpreted in its lexical Stage I meaning, whereas (7(ii)) is suggestive of Stage II, which appears to be an optional variant of (7(i)), showing a grammatical (i.e. a reflexive) meaning. Thus, the construction has undergone semantic change by inviting a bridging Stage II interpretation, while the form seems to have remained unaffected by the change.

(7) Yoruba (Kwa, Niger-Congo; Awolaye 1986: 11)

    won   rí   ara   wọn.
    they saw body their
    (i) 'They saw their bodies.'
    (ii) 'They saw themselves.'

2.2.3 FROM ACTION VERB TO COMPARATIVE MARKER

Cross-linguistically there is a wide range of constructions used to express comparisons between two different items, and there are a number of different comparative concepts that tend to be distinguished. Our concern in this section is with only one of these concepts, namely with the comparative of inequality, or the superior comparative as it has also been called (Stassen 1985).

TABLE 2.4 The main event schemas used for encoding comparative constructions (see Heine 1997a: 112)

Form of schema	Label of schema
X is Y surpasses Z	Action
X is Y at Z	Location
X is Y from Z	Source
X is Y to Z	Goal
X is Y, Z is not Y	Polarity
X and Z, X is Y	Topic

Comparison is a relatively abstract concept, and, as we argue here, expressions of comparison are likely to be historically derived from more concrete meanings via grammaticalization. These meanings have been described in Heine (1997a) in terms of conceptual templates, called event schemas. Cross-linguistically there is only a small set of event schemas that tend to be recruited to grammaticalize comparative constructions; the most common of these schemas are summarized in Table 2.4.

In accordance with these schemas, comparatives are built on concepts such as action, where the standard of comparison is presented by means of an action verb (Action Schema), location (Location Schema), source or ablative (Source Schema), direction or benefactive (Goal Schema), an antonymic relation (Polarity Schema), or in terms of thematic conjuncts (Topic Schema).

In principle, speakers of a given language may select any of the schemas listed in Table 2.4 to develop a new comparative construction; and in many languages, more than one schema has been grammaticalized. It would seem, however, that neigh-bouring linguistic communities are more likely to draw on the same schema than are communities living at some distance from one another. This is suggested by the fact that there are geographically defined macro-areas where a preference for a specific kind of schema can be observed. Table 2.5 summarizes the results of a cross-linguistic survey of these constructions, carried out by Stassen (1985).[7]

Our interest here is with the macro-area of Africa, which exhibits a clear preference pattern: according to Table 2.5, more than half of all African sample languages (56 per cent) have grammaticalized the Action Schema to a comparative construction. But perhaps more significantly, almost two thirds (65 per cent) of all languages of the worldwide sample in Table 2.5 having made use of this schema are spoken in Africa.[8]

As we wish to show now, the context model depicted in Table 2.1 also applies to the grammaticalization of comparatives of inequality. In most cases, the Action

---

[7] The sample of 109 languages has been established on what Stassen (1985) argues is a genetically and areally balanced selection of the world's languages.

[8] Another linguistic area where the Action Schema ('surpass comparative') provides the main source of grammaticalization for comparatives of inequality is mainland Southeast Asia (Ansaldo 1999, 2010).

TABLE 2.5 Event schemas serving as sources for the grammaticalization of comparatives of inequality (primary options only; sample: 109 languages of worldwide distribution; Stassen 1985; Heine 1997a: 128)

Schema	Europe	Asia	Africa	The Americas	Indian/Pacific Ocean	Total
Action	0	4	13	1	2	20
Location	0	4	3	4	1	12
Source	0	18	4	9	1	32
Goal	1	0	3	3	3	10
Polarity	0	0	0	10	10	20
Topic	0	0	0	0	0	0
Opaque schemas*	13	0	0	1	1	15
Total	14	26	23	28	18	109

* 'Opaque schemas' are conceptual sources whose genesis is etymologically not recoverable.

Schema involves a verb meaning 'defeat', 'surpass', or 'pass', but in a few languages there is a verb for 'leave (behind)' instead.

In the !Xun language of southwestern Africa, two of these verb types have been grammaticalized. !Xun (or Ju), formerly known as Northern Khoisan, belongs to the Kx'a family (Heine and Honken 2010); the data presented below are taken from the W2 dialect of !Xun spoken in Ekoka of northern Namibia (Heine and König 2015). The examples presented in (8) illustrate the first pathway, involving the verb ǹ/hūnyā 'leave (behind)', where (8a) represents the lexical and (8b) the grammatical meaning of a marker denoting the standard of comparison. Which of the two meanings is expressed depends on the context in which these verbs are used. The context illustrated in (8a) highlights the spatial meaning of the movement verb ǹ/hūnyā, hence the lexical meaning of Stage I of the context model (Table 2.1) surfaces in this example. In (8b), by contrast, the manner of the action performed is foregrounded; accordingly, the only reasonable interpretation for !Xun speakers is one with reference to Stage III, where the lexical source meaning is backgrounded in favour of the grammatical meaning.

(8)   !Xun (W2 dialect; Kx'a; Heine and König 2005, 2015; König and Heine 2008)
    a. Cālò   má   kē   ǹ|hūnyā   hà   n!āō.
        Calo   TOP   PST   leave   his   house
        'Calo left his house.'

    b. hà   má   ḿ   ǹ|hūnyā   mí.
        N1   TOP   eat   leave   1.SG
        'He eats more than I.'

Essentially the same pathway can be reconstructed for the second verb, !'ālā 'pass (by)'. Thus, in (9a) and (9c) essentially the same applies as in (8a) and (8b), respectively. But in this case there is an intermediate stage, i.e. a bridging stage situation of Stage II, where !'ālā is ambiguous between the lexical meaning (9b (i)) and the grammatical meaning (9b (ii)). The difference between (9b) and (9c) lies in

the context provided by the preceding verb: Whereas *‼’hùᷓm* 'run' can be interpreted in this sentence with reference either to its meaning of spatial movement or to the manner of action, a spatial interpretation is ruled out with the stative verb *n‖ā’à* 'be big'; hence, (9b) is ambiguous while (9c) does not allow for a spatial interpretation: there is only the grammatical concept ('more than') of a comparative marker.

(9) !Xun (W2 dialect; Kx'a; Heine and König 2005, 2015; König and Heine 2008)

    a. mí   má   kē   !’ālā n!āō.
       1.SG TOP PAST pass house
       'I passed by the house.'

    b. !xó   má   ‼’hùᷓm   !’ālā gùmì.
       elephant TOP   run   pass   cattle
       (i) 'The elephant runs, overtaking the cow.'
       (ii) 'An elephant runs better (or faster) than a cow.'

    c. !xó   má   n‖ā’à   !’ālā gùmì.
       elephant TOP be.big pass cattle
       'An elephant is bigger than a cow.'

There is no conventionalized standard marker in !Xun, i.e. grammaticalization has not proceeded beyond Stage III.

    As in sections 2.2.1 and 2.2.2, we hypothesize that there was a change leading from a lexical meaning (Stage I) to a schematic, grammatical meaning without a corresponding change in form: As far as can be ascertained, both items, *!’ālā* 'pass (by)' and *ǹ/hūnyā* 'leave (behind)', are phonologically and morphosyntactically identical, irrespective of whether the lexical source or the grammatical target meanings are expressed. Once again, there is support for the MF hypothesis, according to which grammaticalization was triggered by context-induced semantic change with no corresponding formal change.

2.2.4 A DE-ANDATIVE FUTURE

The final case concerns a pathway of grammaticalization that is widespread in Africa but presumably equally widespread in other parts of the world (see e.g. Bybee et al. 1991; Bybee, Perkins, and Pagliuca 1994; Heine and Kuteva 2002), examples of it can be found in European languages such as English or French. Our example is also taken from the !Xun language of southwestern Africa, but this time from the N1 dialect spoken in southeastern Angola. This dialect uses a weakly grammaticalized future tense, showing the grammaticalization pathway from a lexical construction involving the verb *ú* 'to go', illustrated in (10a), via an ambiguity Stage II situation in (10b), where the grammatical meaning of future tense is foregrounded but the lexical meaning is still available. This is the situation that obtains with most kinds of verbs serving as complements of *ú* 'go' in this dialect. However, when the meaning of the complement verb is semantically incompatible with that of *ú* 'go', then the lexical source meaning is ruled out. This is the case in (10c), where the spatial deixis of the complement verb *tcí* 'come' is incompatible with that of *ú* 'go'; in such a