

WORLD LEXICON OF GRAMMATICALIZATION

Bernd Heine ♦ Tania Kuteva

This page intentionally left blank

World Lexicon of Grammaticalization

While the comparative method is concerned with regularities in phonological change, grammaticalization theory deals with regularities of grammatical change. In an A–Z format, this book summarizes the most salient generalizations that have been made on the unidirectional change of grammatical forms and constructions. The product of ten years of research, *World Lexicon of Grammaticalization* provides the reader with the tools to discover how different grammatical meanings can be related to one another in a principled way, how such issues as polysemy and heterosemy are dealt with, and why certain linguistic forms have simultaneous lexical and grammatical functions. It covers several hundred grammaticalization processes, in each case offering definitions of lexical concepts, suitable examples from a variety of languages, and references to the relevant research literature; appendixes organized by source and target concepts allow for flexible use. The findings delineated in the book are relevant to students of language across theoretical boundaries.

The author of thirty-two books, Bernd Heine is Professor at the Institute for African Studies at the University of Cologne.

Tania Kuteva is Professor of English Linguistics at the University of Düsseldorf.

World Lexicon of Grammaticalization

Bernd Heine

University of Cologne

Tania Kuteva

University of Düsseldorf



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Bernd Heine and Tania Kuteva 2004

First published in printed format 2002

ISBN 0-511-03273-0 eBook (Adobe Reader)

ISBN 0-521-80339-X hardback

ISBN 0-521-00597-3 paperback

Contents

Acknowledgments

vii

Abbreviations and Symbols

ix

1

Introduction

1

2

Grammatical Concepts Used in This Work

15

3

Source–Target Lexicon

27

Appendix 1: Source–Target List

317

Appendix 2: Target–Source List

327

Appendix 3: List of Languages

337

References

351

Acknowledgments

World Lexicon of Grammaticalization is based on research carried out between 1991 and 2000 at the University of Cologne and the Center for Advanced Study in the Behavioral Sciences, Stanford, California. A number of people have contributed to this work in various ways, in particular the following: Sasha Aikhenvald, Johan van der Auwera, Walter Bisang, Jürgen Broschart, Joan Bybee, Jean-Pierre Caprile, Hilary Chappell, Ulrike Claudi, Gerrit Dimmendaal, Bob Dixon, Spike Gildea, Tom Givón, Colette Grinevald, Tom Güldemann, John Haiman, Martin Haspelmath, Ingo Heine, Paul Hopper, Raimund Kastenholtz, Roland Kießling, Christa Kilian-Hatz, Alan King, Christa König, Sylvia Kutscher, Beth Levin, Thomas Müller-Bardey, John Newman, Helma Pasch, Alain Peyraube, Mechthild Reh, Susanne Romaine, Mathias Schladt, Chao Fen Sun, Bernardus Suwute, Kossi Tossou, Elizabeth Traugott, Holger Tröbs, Erhard Voeltz, Manfred Weber, and Elizabeth Weise. Finally, we wish to thank ten anonymous readers of Cambridge University Press for their constructive criticism and a wealth of additional information they contributed to this volume.¹

Research on which this volume is based was generously sponsored by the Deutsche Forschungsgemeinschaft (German Research Society), to which we wish to express our deeply felt gratitude. The first-named author also wishes to thank the Center for Advanced Study in the Behavioral Sciences, Stanford University, for offering him an academic home during the final phase of work on this book. The second-named author is grateful to the Osterlits Research Laboratory, CNRS (Centre National de la Recherche Scientifique), Paris, France, for its support.

¹ We have marked such examples with the phrase “anonymous reader.”

Abbreviations and Symbols

ABL	ablative	CAU	cause
ABS	absolutive	CAUS	causative
ACC	accusative	CD	Dutch-based creole
ACTU	actuality	CE	English-based creole
ADE	adessive	CF	French-based creole
ADJV	adjectivizer	CL	noun class
ADV	adverb	CLASS	classifier
ADVL	adverbial	CM	class marker
ADVR	adverbializer	CN	class noun
AGR	agreement	COM	comitative
ALL	allative	COMP	complementizer
ALM	almost	COMPAR	comparative
AOP	aorist participle	CON	continuous
AOR	aorist	COND	conditional
APPL	applicative	CONJ	conjunction
ART	article	CONN	connective suffix
ASP	aspect	CONP	connective particle
ASS	assertion particle	CONT	continuative
AUX	auxiliary	CONV	converb
AVER	avertive	COP	copula
BEN	benefactive	CP	Portuguese-based creole
C	common gender; creole when in reference to specific languages	CPL	completive
C1, C2, etc.	noun class 1, 2, etc.	CRS	currently relevant state
CA	Arabic-based creole	CS	Spanish-based creole

DAT	dative	IMPFV	imperfective
DEB	debitive	INAN	inanimate
DEC	declarative	INCL	inclusive
DEF	definite	INCPL	incompletive
DEICT	deictic marker	IND	indicative
DEM	demonstrative	INDEF	indefinite
DET	determinator	INE	inessive
DIM	diminutive	INF	infinitive
DIR	directional	INSTR	instrument
DISC	discontinuous	INSTRN	instrumental
DISTAL	distal (demonstrative)	INT	intentional
DU	dualis	INTER	interrogative marker
DUR	durative	INTR	intransitive
EMPH	emphatic	INTJ	interjection
EMPTY	empty marker	IRR	irrealis
ERG	ergative	IS	intransitive final suffix
EXCL	exclusive		
F	feminine gender	JUNC	juncture
FEM	female		
FACT	factitive	LIG	ligature
FOC	focus	LNK	linking vowel
FREQ	frequentative	LOC	locative
FUT	future	LOG	logophoric
GEN	genitive		
GER	gerund	M	masculine gender
GL	goal	MO	motion verb class
HAB	habitual		
HON	honorific	N	noun
HORT	hortative	NAR	narrative
HUM	human	NEG	negation
		NEUT	neuter gender
IDENT	identifier	NFUT	near future
IDEO	ideophone	NOM	nominative
ILL	illative	NOMIN	nominalizer
IMC	imperfective converb	NONPAST	nonpast
IMP	imperative	NP	noun phrase
IMPERF	imperfect	NPERF	near perfect
		NPL	nonplural marker

OBJ	object	REAS	reason
OBL	oblique	REC	reciprocal
OPT	optative	REFL	reflexive
P	pidgin, in reference to specific languages	REL	relative (clause marker)
PA	Arabic-based pidgin	RES	restrictive
PART	particle	RESULT	resultative
PARTCP	participle	RM	relator, relation marker
PARTV	partitive	SBST	substantivizer
PASS	passive	SG	singular
PAST	past	SRDIR	superdirective marker
PE	English-based pidgin	SREL	superrelative marker
PERF	perfect	SRESS	superessive marker
PFV	perfective	SS	same subject marker
PL	plural	STATS	subject of a stative verb
PLU	pluperfect	SUB	subordinator
PM	participial marker	SUBEL	subelative marker
POESS	postessive	SUBJ	subject
POSS	possessive	SUBJUNCT	subjunctive
POST	postposition	SUF	suffix
POT	potential	TAM	marker of tense, aspect, or modality
PRED	predicate marker	TERM	terminative
PREP	preposition	TNS	tense
PRES	present	TOP	topic
PROG	progressive	TR	transitive
PROH	prohibitive	TRI	trial
PRON	pronoun	VEN	venitive
PROX	proximative	VINC	incremental
PROXIM	proximal (demonstrative)	VN	vowel on verbs verbal noun
PERS	person		
PST	participle of state		
PURP	purpose		
PX	proximity marker		
Q	interrogative		
QUOT	quotative		
R	relational suffix		
REAL	realis		

1	first person
2	second person
3	third person
I	juncture I
II	juncture II
=	clitic boundary
?	morpheme of unknown meaning
*	reconstructed item
()	interlinear glosses tentatively volunteered by the authors

Introduction

Over the course of the last three decades, a wealth of data has been published on the origin and development of grammatical forms. The main purpose of the present work is to make this wealth accessible to a wider readership. To this end, over 400 processes relating to the evolution of grammatical categories are discussed, using data from roughly 500 different languages. (See Appendix 3 for a list of languages figuring in this book.)

The readership we have in mind for this book includes first of all linguists. Grammaticalization theory, which is the framework adopted here (see §1.1), is concerned with language use across space and time; hence the findings presented may be of help for diachronic reconstruction, especially in areas where other tools available to the historical linguist, such as the comparative method and internal reconstruction, do not yield appropriate results. The descriptive linguist will find information, for example, on how and why different grammatical meanings can be related to one another in a principled way (i.e., on how to deal with issues like polysemy and heterosemy), on why there are some regular correspondences between grammatical forms and the meanings expressed by them, or on why certain linguistic forms have simultaneously lexical and grammatical functions. Anthropologists, sociologists, and psychologists may discover that the kind of human behavior held responsible for the evolution of grammatical forms is not all that different from the kind of behavior they observe in their own fields of study.

What distinguishes this work from relevant monographs on grammaticalization theory (e.g., Lehmann 1982; Heine and Reh 1984; Heine, Claudi, and Hünemeyer 1991; Traugott and Heine 1991a, 1991b; Hopper and Traugott 1993; Bybee, Perkins, and Pagliuca 1994; Pagliuca 1994; Heine 1997b; Ramat and Hopper 1998) is its conception as a reference work. Accordingly, an attempt was made to collect many data from as many different languages as possible and to avoid theoretical biases – as far as this is possible and feasible.

1.1 Grammaticalization Theory

Grammaticalization is defined as the development from lexical to grammatical forms¹ and from grammatical to even more grammatical forms. Since the development of grammatical forms is not independent of the constructions to which they belong, the study of grammaticalization is also concerned with constructions and with even larger discourse segments.

In accordance with this definition, grammaticalization theory is concerned with the genesis and development of grammatical forms. Its primary goal is to describe how grammatical forms and constructions arise and develop through space and time, and to explain why they are structured the way they are.² Technically, grammaticalization involves four main interrelated mechanisms.

- (a) desemanticization (or “semantic bleaching”) – loss in meaning content,
- (b) extension (or context generalization) – use in new contexts,
- (c) decategorialization – loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms, and
- (d) erosion (or “phonetic reduction”) – loss in phonetic substance.

While three of these mechanisms involve a loss in properties, there are also gains. In the same way that linguistic items undergoing grammaticalization lose in semantic, morphosyntactic, and phonetic substance, they also gain in properties characteristic of their uses in new contexts. Grammaticalization requires specific contexts to take place, and it can be, and has been, described as a product of context-induced reinterpretation. Accordingly, context is a crucial factor in shaping the structure of grammatical forms – to the extent that they may express meanings that cannot immediately be derived from their respective source forms.

It has been argued that grammaticalization is not a distinct process, since the four mechanisms can be observed to be at work also in other kinds of linguistic change (Newmeyer 1998: 248ff.).³ There are a couple of reasons why we think that such a position is not justified. First, the main task of grammaticalization theory is to explain why grammatical forms and constructions are structured the way they are, and these four

¹ The term “grammatical forms,” or “grams,” roughly corresponds to what is also referred to as “functional categories.”

² Newmeyer (1998: 240) raises doubts about whether we are really dealing with a theory here, and he rightly observes that much of the relevant literature on this subject is not very helpful on deciding this issue.

³ Newmeyer (1998: 260) refers to desemanticization as “appropriate semantic change,” to decategorialization as “downgrading analysis,” and to erosion as “phonetic reduction.”

mechanisms, as opposed to many other conceivable mechanisms, have been found to be relevant to achieve such explanations. Thus, irrespective of how one wishes to define a “distinct process,” one is led to conclude that these mechanisms are part of one and the same explanatory framework.

Second, grammaticalization, as conceived here, is above all a semantic process. This process is context dependent, and grammaticalization can therefore be described in terms of context-induced reinterpretation. Not every reinterpretation leads to the rise of grammatical meanings. Rather, it is only when forms for concrete (e.g., lexical) meanings are used to also express more abstract (grammatical) meanings that grammatical forms emerge; for example, when a form used for a visible object (e.g., the body part ‘back’) is used also to refer to a nonvisible item (the spatial notion ‘behind’), or a form used for an action (‘go to’) is used also to refer to a grammatical notion (future tense). On account of its specific directionality, context-induced reinterpretation has been described in terms of metaphorical transfer, leading, for example, from the domain of concrete objects to that of space, from space to time, from (“real-world”) space to discourse space, and so on.

Desemanticization thus results from the use of forms for concrete meanings that are reinterpreted in specific contexts as more abstract, grammatical meanings. Having acquired grammatical meanings, these forms tend to become increasingly divergent from their old uses: they lose in categorial properties characteristic of their old uses, hence undergoing decategorialization, and they tend to be used more frequently, to become more predictable in their occurrence, and, consequently, to lose in phonetic substance. Thus, the four mechanisms are not independent of one another; rather, desemanticization precedes and is immediately responsible for decategorialization and erosion. There are a few cases where it has not yet been possible to establish that decategorialization really followed desemanticization in time, and we do not wish to exclude the possibility that in such cases the two may have occurred simultaneously. However, such cases appear to be exceptional: new grammatical meanings arise, and it usually takes quite some time before any corresponding morphological, syntactic, and/or phonetic changes can be observed. In many languages, prepositions unambiguously serving a grammatical function still have the morphosyntactic structure of their earlier uses as adverbial phrases (cf. English *by means of*, *in front of*, *with respect to*) or verbal phrases (cf. Chinese ZAI ‘(to be) at’; Alain Peyraube, personal communication), and tense or aspect auxiliaries may still behave morphosyntactically largely like lexical verbs even if they have lost their lexical semantics and serve exclusively as functional categories (cf. English *be*

going to, *used to*, *keep (doing)*, etc.). To conclude, there is evidence to suggest that grammaticalization can be defined as a distinct process.

It is sometimes assumed that grammaticalization invariably involves lexical categories; that is, that it is confined to the development from lexical to grammatical forms. This view tends to ignore that such cases account for only part of what falls under the rubric of grammaticalization. Equally commonly, as we will see in the course of this work, items already part of the inventory of grammatical forms give rise to more strongly grammaticalized items. Prepositions often develop into conjunctions, temporal conjunctions tend to give rise to causal or concessive conjunctions, demonstrative determiners develop into definite articles or relative clause markers, verbal perfect inflections may become past tense markers, and so forth – all developments that take place within the domain of functional categories. Such developments are distinguished mainly from developments involving lexical categories by the difficulty of identifying and reconstructing them.

Grammaticalization is a unidirectional process; that is, it leads from less grammatical to more grammatical forms and constructions. However, this process is not without exceptions: a number of examples contradicting the unidirectionality principle have been found (see, e.g., Joseph and Janda 1988; Campbell 1991; Ramat 1992; Frajzyngier 1996; and especially Newmeyer 1998: 260ff.). Yet, as acknowledged by most of the scholars who have identified exceptional cases, such examples are few compared to the large number of cases that conform to the principle⁴ (cf. Haspelmath 1999, 2000: 249). Furthermore, they can frequently be accounted for with reference to alternative forces,⁵ and finally, no instances of “complete reversals of grammaticalization” have been discovered so far (cf. Newmeyer 1998: 263).

Grammaticalization begins with concrete, lexical forms and constructions and ideally ends in zero – that is, grammatical forms increasingly

⁴ Cf., e.g., Harris and Campbell (1995: 338), who summarize this situation thus: “there is a strong tendency for grammaticalization to proceed in one direction, though it is not strictly unidirectional.” Similarly, Joseph and Janda (1988: 198–200) observe that cases of demorphologization, a process that would contradict the unidirectionality principle, are rare and not seldom controversial. Finally, Newmeyer (1998: 275–6, 278) observes that cases conforming to the unidirectionality principle (“downgradings”) “have occurred at least ten times as often as upgradings,” and he concludes, “I suspect that, for whatever reason, there is a *general* directionality to the semantic changes observed in grammaticalization” (emphasis in original).

⁵ Such forces may be morphophonological or morphosyntactic in nature, but they may as well relate to specific sociocultural factors. Burridge (1995) discusses an example of reversed directionality in Pennsylvania German, where a modal auxiliary developed into a lexical verb, *wotte* ‘wish’. As Burridge shows, one factor contributing to this development can be found in the particular Mennonite religious principles held by the speakers of Pennsylvania German.

lose in semantic and phonetic content and, in the end, they may be replaced by new forms; grammaticalization has therefore been described as a cyclical process (Givón 1979a; Heine and Reh 1984).⁶ While there is some evidence to support this assumption, we have to be aware that, first, a grammaticalization process can stop at any point of development and, second, “worn-out” grammatical forms are not necessarily replaced by new forms. Thus, the metaphor of a grammatical cycle, though useful in certain cases, should not be generalized since it often does not apply for some reason or other.

In a number of works, grammaticalization is described as a process that involves the reanalysis of grammatical categories.⁷ Other authors have argued that there is no necessary relationship between grammaticalization and reanalysis (see especially Haspelmath 1998). In fact, reanalysis has been defined in a number of different ways (cf. Langacker 1977; Heine and Reh 1984; Harris and Campbell 1995: 61–96; Haspelmath 1998; Newmeyer 1998: 241–51). Whether grammaticalization involves reanalysis has turned out to be essentially a theory-dependent issue. To avoid any further confusion on this issue, we prefer to exclude “reanalysis” from our terminology of grammaticalization theory.

1.2 Problems

Grammaticalization is a complex subject matter; it relates in much the same way to diachronic and synchronic linguistics as to semantics, syntax, and morphology, and it is rooted in cognition and pragmatics. Obviously, an endeavor such as that found here is an ambitious one – one that has to take care of a wide range of problems. In this section we deal with the most serious of these problems in turn.

The findings presented in this work are meant to highlight processes of human behavior that can be observed across cultures; yet, these findings are based on data from hardly more than one-tenth of the world’s languages. One may therefore wonder what justification there is to call this work a “world lexicon.” Our main reason is this: underlying human behavior there appears to be a strategy of linguistic processing whereby more abstract functions are expressed in terms of forms for concrete concepts. We expect, for example, that in some unknown language there are

⁶ Givón (1979a: 209) proposed the unidirectional cycle in (i), where the end point (Zero) marks the beginning of a new cycle again leading from Discourse to Zero:

(i) Discourse > Syntax > Morphology > Morphophonemics > Zero.

⁷ Newmeyer (1998: 238), for example, argues, “The standard definition of grammaticalization incorporates the notion of reanalysis; no definition that does not do so seems particularly useful.”

ways of expressing temporal concepts in terms of spatial ones, spatial relations in terms of forms for concrete concepts (such as body parts or salient landmarks), aspectual contours of events in terms of forms for actions and motions, or functions concerning the organization of texts in terms of linguistic forms for spatial or temporal deixis. Languages differ considerably in the way and the extent to which this strategy has given rise to grammaticalized constructions; nevertheless, we expect the effects of this strategy to be essentially the same across languages, including languages that are still undocumented.

Throughout this work we are concerned with the relation between two kinds of concepts, which we refer to as the “source” and “target” entities of grammaticalization. We convey the impression in this account that there is always a unidirectional development leading from one distinct entity to another entity. But this is not only a simplified account; it is also at variance with much of what we have argued for elsewhere, namely that, rather than being a development in discrete steps, grammaticalization must be described as a continuous or, more precisely, as a chainlike development (Heine 1992). To achieve the goal of having a treatment of grammaticalization processes in the form of a lexicon, we were forced to reduce continuous, chainlike structures to two salient uses of forms, viz., source and target uses.

Most of the over 400 grammaticalization processes discussed in this book are based on fairly reliable reconstruction work, but in some cases the evidence available is not yet satisfactory. We have pointed out such cases under the relevant entry.

A number of developments leading to the evolution of grammatical categories do not involve linguistic units like words or morphemes (Heine 1993; Bybee et al. 1994; Bisang 1998a); rather, they concern more complex conceptual entities, such as phrases, whole propositions, or even larger constructions. For example, the temporal conjunction *taátenu* ‘then’ of Kxoe, a Central Khoisan language of Namibia, is historically a clause meaning ‘when it is like that’ (see (1)).

- (1) *ta- á- te- nu xavána //é kúùn-à- tè*
 be:thus-JUNC-PRES-when again 1:M:PL go- JUNC-PRES
 ‘Then we went again. . . .’

A much better known example concerns the evolution of aspect and tense categories, where two or more different linguistic forms may simultaneously be involved: an auxiliary (e.g., *be* or *have*), a nonfinite marker (e.g., an infinitival, participial, or gerundival marker), and perhaps also a locative marker. Tense and aspect constructions in a number of languages worldwide not uncommonly involve three distinct morphological

elements, the English future marker *be going to* being a paradigm example. Another European example is the Latin verb *habere* ‘to have’, which in the Romance languages has given rise to perfect markers on the one hand and to future markers on the other. What accounts for this divergent development? The verb *habere* was not itself grammaticalized; rather grammaticalization involved entire periphrastic constructions, or event schemata: the construction *habere* + perfect passive participle gave rise to perfect expressions, while *habere* + infinitive periphrasis was responsible for the development of future constructions. In a lexicon project like the present one, such propositional structures had to be reduced to the salient segments of the constructions concerned, such as the *habere*-markers figuring in the expression of future tenses in Romance languages.

A related problem that we encountered concerns what one may call “complex grammaticalization”: a more complex linguistic structure can assume a grammatical function without involving the grammaticalization of any particular item figuring in this structure. Take (1) again: which of the various items figuring in the Kxoe word *taátenu* should be held responsible for the relevant grammaticalization? The most obvious answer would be that, rather than any particular item, the structure as a whole is responsible. In a treatment of the kind attempted here, however, which rests on the assumption that there is essentially a one-to-one correspondence between source and target, such an answer is not entirely satisfactory. What exactly should the lexicon entry be that takes care of this grammaticalization? Or take the following example: one widespread way of developing expressions for the grammatical concept of a comparative of inequality is to juxtapose two propositions that are in a polar contrast – one expresses the standard of comparison and the other the comparative notion. This opposition may be either antonymic, as in (2), or marked by the distinction of positive versus negative, as in (3).

Cayapo (Stassen 1985: 184)

- (2) *Gan ga prik, bubanne ba i pri.*
 you you big but I I small
 ‘You are bigger than I am.’

Abipon (Stassen 1985: 184)

- (3) *Negetink chik naâ, oagan nihirenak la naâ.*
 dog not bad yet tiger already bad
 ‘A tiger is more ferocious (lit.: ‘bad’) than a dog.’

What is grammaticalized in such constructions is not a specific element but rather some propositional relation, viz., *be big* versus *be small*, or *be*

bad versus *not be bad*. In a treatment like this book, which is concerned with segmentable linguistic forms, functions expressed by means of pragmatic or syntactic relations between forms without involving morphological segments of necessity had to be excluded.

The sentence in (3) raises another question: At which point can we say that grammaticalization has been concluded? Can we really say that (2) and (3) are suggestive of a completed process of grammaticalization, or do they merely represent contextually induced interpretations that are irrelevant for the grammatical structures of the languages concerned? A number of tests have been proposed in grammaticalization theory to deal with this question; frequently, however, the information available on a given language is not sufficient to allow for a successful application of these tests. In such cases we have decided to adopt the solution proposed by the author(s) dealing with that language.

In some cases we decided to rely on comparative findings to determine whether a grammaticalization process has been concluded. For example, one of our entries has the form ONE > INDEFINITE, according to which the cardinal numeral for 'one' may grammaticalize to indefinite articles. Now, it has been argued, for languages like English (*a(n)*) or German (*ein*), for example, that the two, numeral and indefinite article, are the same, their difference being due to contextual or other factors; that is, that the relevant entry is not an instance of grammaticalization. That the two meanings are in fact different is suggested by comparative observations. Thus, there are languages where a given linguistic item serves as an indefinite marker but not as a numeral, and, conversely, there are many languages where a given item denotes the numeral 'one' but not indefinite reference. We take such observations as evidence that ONE and INDEFINITE are in fact different concepts, even if in some languages the same or a similar word is used for both.

Another problem concerns the directionality of grammaticalization and how to achieve historical reconstruction. How do we know that INDEFINITE is historically derived from ONE rather than the other way around? In this case, there is diachronic evidence to give an answer: in some languages, including a number of European ones, there is a marker that is used for both the numeral 'one' and the indefinite article, and by using historical records it is possible to establish that at some earlier stage in the development of these languages the item only served as the numeral expression before its use was extended to also designate indefinite reference. Now, since grammaticalization is essentially unidirectional, we are led to assume that in languages where no historical records are available the evolution was the same.

Even in the absence of historical documents it is possible to reconstruct directionality of change by using the mechanisms sketched in the preceding section. For example, decategorialization has the effect that the element concerned loses in morphosyntactic properties characteristic of its less grammaticalized (e.g., lexical) source, such as the ability to take modifiers or inflections, and it shifts from a category having many members (e.g., an open class) to a category having only few members (a closed class). Erosion again means that that element tends to become shorter and/or phonetically less complex, to lose the ability to receive distinct stress or tone, and so on. Thus, if we find two different uses of a given element, or two etymologically related elements, where one shows the effects of decategorialization and erosion whereas the other does not, then we can argue that the latter is less grammaticalized and then reconstruct a directionality from the latter to the former, rather than the other way around. Even if we had no previous knowledge of the history of English we could nonetheless establish that the indefinite article *a(n)* is a later development form of the numeral *one*, rather than the reverse, since the article exhibits a number of effects of decategorialization and erosion while the numeral does not. In this text we use this kind of evidence for reconstruction in addition to any kind of historical evidence that may be available.

Grammaticalization does not occur in a vacuum, and other forces also shape the evolution of grammatical forms, language contact being one. The rise of a new grammatical expression may be the result of grammaticalization, but it may also be due to the influence of another language. The question of whether, or to what extent, a given development is from language-internal as opposed to language-external factors can frequently not be answered satisfactorily. Recent studies suggest that both are often simultaneously involved.

These observations led us to the question of whether any restriction in the kind of linguistic transmission should be imposed when selecting the data to present in this volume. For example, should instances of grammaticalization that clearly occurred due to borrowing be excluded? Should we separate such cases from instances of grammaticalization that have to do with continuous transmission within a given language?

A perhaps related issue concerns pidgins and creoles, which are a gold mine for students of grammaticalization, and throughout the 1990s a wealth of publications appeared demonstrating the relevance of grammaticalization theory to the study of these languages (see especially Baker and Syea 1996). With the rise of pidgins and creoles, the question again arises as to whether we are dealing with “natural” forms of transmission

and, if yes, whether grammaticalization processes behave the same way whether they have taken place, for example, between earlier and later forms of British English or between British English and Krio CE or Tok Pisin PE. The policy adopted here is to take all these kinds of data into account, at least as far as they are in accordance with principles of grammaticalization observed in “natural” language transmission. More recent research suggests that grammaticalization in pidgins and creoles does not behave essentially differently from that found in other languages. The reader is in a position to identify instances of borrowing or pidginization, or creolization, on the basis of the exemplification provided in this book.⁸

The terminology used to refer to grammatical categories differs from one author to another and from one language to another. Although we have tried to standardize terms, in many cases, this turned out to be impossible because of insufficient information. It is therefore to be expected that, in accordance with the conventions adopted by the relevant authors, one and the same grammatical function may be referred to by entirely different labels, both within a given language and across languages.

The quality of the data provided in this work crucially depends on the kind of information contained in the published sources that we were able to consult. Frequently it turned out that the information was not satisfactory. For example, when dealing with a verb as the source for a certain grammatical category, it is not enough to consider the lexical semantics of that verb; which grammaticalization it undergoes may depend entirely on its valency. In Southern Sotho, a Bantu language of Lesotho and South Africa, we find, among others, instances of grammaticalization like those presented in (4).

Southern Sotho (Bantu, Niger-Congo; Doke and Mofokeng [1957] 1985)

(4) Verbal source	Grammatical form
-ea ‘go (to)’	-ea- immediate future tense
-tla ‘come (to)’	-tla- future tense
-tsoa ‘come from’	-tsoa- immediate past tense

These examples suggest that it is not the deictic semantics of ‘come’ or ‘go’ that can be held responsible for the particular functions the result-

⁸ Pidgin (P) and creole (C) examples are marked by adding abbreviated labels after the language name. For example, “CE” stands for “English-based creole” (see Abbreviations). Note that the classification underlying this usage is a crude one, since terms like “English-based,” “Portuguese-based,” etc. are not unproblematic, and the boundary between pidgins and creole languages is not seldom fuzzy.

ing grammatical categories assume; rather, it is the kind of complements they take that determines their path of grammaticalization. If the verb takes an allative/goal complement, as in the case of Southern Sotho *-ea* and *-tla*, then the resulting function is future; if the verb takes an ablative/source complement, as in the case of *-tsoa*, then the result is a perfect or near past category (see Bybee, Pagliuca, and Perkins 1991). Unfortunately, most published sources that we were able to consult do not provide information of this kind. Due to such factors, our documentation must remain fragmentary in many cases.

This book is based on hypotheses on diachronic development. In a number of cases, these hypotheses have been adopted from the sources cited, but in others they were not contained in the relevant sources. For example, if in a given grammar the author states that the adverb 'behind' is "homophonous" with or "resembles" the noun 'back', or "may be historically related" to the noun 'back', then the assumption made here on the basis of a larger corpus of cross-linguistic data is that we are dealing with an instance of the grammaticalization of a body part noun to a locative adverb. The reader is therefore reminded that a given author whose work is cited as evidence for some reconstruction is not necessarily to be held responsible for the relevant reconstruction, such responsibility being entirely ours.

Perhaps the most crucial problem we were confronted with concerns directionality. As some recent works suggest, there are exceptions to the unidirectionality principle,⁹ and we certainly do not exclude the possibility that some of the reconstructions presented allow for an alternative analysis. Still, such cases are likely to be statistically insignificant: the tense markers listed in (4) can be assumed to be derived from verbs of motion, while we know of no language where there is compelling evidence that a verb meaning 'go' or 'come' is historically derived from a tense marker. Yet, the question of directionality is one that needs more attention in future work on grammaticalization.

This lexicon differs in a number of ways from Heine et al. (1993). Above all, whereas the discussion in Heine et al. (1993) was concerned with both the meaning and the morphosyntax of linguistic forms, we confine ourselves here to the analysis of grammatical "concepts." Accordingly, no reference is made to the word or morpheme status of the items undergoing grammaticalization, unless there are specific reasons to do so.

All instances of conceptual shift are illustrated with examples from different languages whenever appropriate data were available. In a number

⁹ A number of exceptions to the unidirectionality principle have been pointed out in recent works (see Newmeyer 1998 for a detailed discussion).

of cases, however, such data could not be found, and we had to rely on hypotheses put forward by other authors. In such cases, the reader is referred to the bibliographical references added for further information.

Another problem we were constantly confronted with was the following: how many examples should be adduced to illustrate a given instance of grammaticalization? There was no problem in cases where only a handful or even fewer examples were found for a certain path of grammaticalization. But for the many cases where the number of possible examples turned out to be exceedingly high, we adopted the policy of reducing exemplification to cases that illustrate both the genetic and areal distribution and the contextual diversity associated with the relevant grammaticalization process. Accordingly, the examples presented here do not necessarily reflect the entire mass of evidence that we were able to assemble. Nevertheless, in the vast majority of cases the amount of exemplification presented immediately correlates with the present state of our knowledge; that is, a grammaticalization process that is amply documented tends to receive a more extensive treatment than one where only a handful of examples have been found so far.

We noted earlier in this chapter that in recent years quite a number of studies have appeared reporting on new processes of grammaticalization (see especially Heine et al. 1991). However, the data presented in this volume constitute but a fraction of all instances of presumed or actual grammaticalization that we were confronted with. There were two reasons for reducing the vast amount of reported processes. First, to strengthen the hypothesis that we are really dealing with cross-linguistic regularities of grammatical evolution, we concentrated on cases where examples from more than one language family were available, even if in the end we decided to also include a number of less widespread instances of grammaticalization whenever there were specific reasons to do so. Second, we eliminated those cases where we were not convinced that the data allowed for fairly reliable reconstruction work. Not all of the processes that have been proposed in the course of the last three decades are substantiated by appropriate empirical evidence. In fact, deciding on whether there is "appropriate empirical evidence" turned out to be one of the major problems we faced when working on this volume.

Finally, we were also confronted with a problem that most lexicographers are confronted with: the closer one gets to completing a lexicon the more one tends to become convinced that one is dealing with an open-ended project and that one is still far from having exhausted the subject matter. But this problem is perhaps even more serious here than in conventional works on lexicography since grammaticalization is a young and rapidly expanding field of research. The reader should therefore be aware

that what is covered in this book might represent merely the tip of the iceberg of what future generations of researchers might discover on this phenomenon.

1.3 Conventions

For a better understanding of the Source-Target lexicon, the following conventions should be borne in mind:

- (a) Entries contain two kinds of information. The first consists of data from different languages, especially from languages that, to our knowledge, are genetically “unrelated.” The second concerns our analysis of this information, that is, our classification and diachronic interpretation of these data. To distinguish these two, all information relating to the latter is printed in small capital letters. Items printed in small capitals each stand for a cluster of closely related meanings (or functions) that we assume to be suggestive of a cross-culturally relatively stable concept. The term “concept” is used as a pre-theoretical notion; no claim is made, for example, that the concepts presented are semantic primitives of any kind or that the label used to refer to a particular concept is suggestive of a prototypical manifestation of that concept.
- (b) To save space, the concept labels are kept as short as possible. Thus, instead of writing “ablative case marker,” or “ablative gram,” we simply use the label “ABLATIVE.”
- (c) Details on the cluster of meanings subsumed under the relevant concept label are provided in parentheses whenever this was felt to be desirable; this parenthetical information is maximally of three kinds. First, it may contain the concept that taxonomically includes the one preceding the parentheses. For example, the concept HEAD has the gloss ‘body part’ following it in parentheses, or ONE has ‘numeral’ added in parentheses. Such parenthetical information is presented in the index of grammatical concepts in Chapter 2. Whenever concepts are involved that do not figure in this index – that is, when lexical concepts are involved – this information is added in the main text (e.g., HEAD (body part)). Second, typical glosses are provided that one might expect to figure in English expressions for the given concept. Third, wherever necessary, these glosses are followed by further descriptive details on the relevant concept.
- (d) At the end of an entry, there may be more general comments relating to the nature of the grammaticalization process in question.

- (e) In the course of our work we were confronted with a number of orthographical issues and problems. As far as this was feasible, we rendered linguistic data in their original form, typically in the standard form used for the language (at least as far as the standard form is based on Roman script). For example, as one might expect, we are using the tilde to mark nasalized vowels (or consonants). There are, however, regional conventions that we also had to take into account. In Nama (of the Khoisan family), nasalized vowels are not marked by a tilde but rather by a circumflex (accent mark: ^); in the standard orthography of Kikuyu and Kamba there is again a tilde, but it does not mark nasalization but rather open vowels.
- (f) Wherever possible we present examples with interlinear glosses. Those printed in parentheses stand for glosses (and in a few cases also translations) that are not in the original examples; for these we take full responsibility. In some cases there were no glosses in the original nor were we able to find appropriate glosses ourselves. We nonetheless decided to include such examples, hoping that the reader interested in more details will consult the bibliographical references cited.
- (g) Our goal is to illustrate all examples with text material, where one text piece, marked by (a), would present the source use and a second text piece, marked by (b), the target use of the item. In most cases, however, no appropriate text material was available, and we had to be satisfied with presenting sentence examples or phrases, or with simply providing a target use without a corresponding source use. We hope that such inconsistency, which is inherent in comparative projects such as this one, is not an obstacle to the use of this work.

Grammatical Concepts Used in This Work

The following list is a classification of the grammatical concepts (or functions) figuring in this work, where the term concept is used in a pre-theoretical sense.¹ Since we will be dealing with concepts, terms such as *ABLATIVE* or *COMPLEMENTIZER* stand for semantic-functional, rather than morphological or syntactic, categories. No attempt is made here to trace a boundary between “grammatical concepts” and nongrammatical or “lexical concepts.” If one finds concepts such as *ONLY* or *TOGETHER*, for example, which one might not be inclined to treat as grammatical concepts, then we simply wish to say that these items exhibit more grammatical properties, or fewer lexical properties, than the concepts from which they are historically derived. Such properties relate in particular to the productivity, applicability to various contexts, and syntactic and paradigmatic status of the items. For example, grammatical forms are closed-class items, and whenever we found that a given concept is regularly derived from some closed-class item we decided to consider it a candidate for inclusion. Both *ONLY* and *TOGETHER* have the numeral *ONE* as one of their historical sources, and although numerals have a fairly large membership in some languages, they normally can be described as closed-class paradigms; hence we decided to tentatively include items such as these two in our treatment.

Furthermore, the characterizations and taxonomic labels that we propose are not intended to be definitions of the concepts; rather, they are meant to assist the reader in narrowing down the range of meanings that a given grammatical marker may convey (see, e.g., Bybee et al. 1994 for more details); in a number of cases, such characterizations consist of nothing but English translational equivalents – a procedure that certainly is far from satisfactory.

¹ We wish to express our gratitude to Beth Levin for many critical comments on the terms presented in this chapter.

In addition to the concept label, the reader will find additional labels in parentheses referring to taxonomically superordinate, more inclusive categories. Since a given concept may belong to more than one more inclusive category, more than one term may appear in parentheses. For example, the entry **ACROSS** (**SPATIAL**, **CASE**) stands for a concept **ACROSS**, which belongs to the concepts used for introducing nominal participants (**CASE**); at the same time, it is also part of the more inclusive category of **SPATIAL** concepts. Rather than reflecting a taxonomy of grammatical concepts, this parenthetical information is simply meant to provide more information on the uses of the primary concept. Yet, there will be cases where the reader may be puzzled as to the exact meaning of a given concept label; in such cases, we refer to the language data presented in the Source-Target lexicon (Chapter 3), where more information on the use of these labels can be found.

Many of the terms presented here are used by other authors to refer to somewhat different, or even to entirely different, concepts. Wherever we are aware of such contrasting uses we point them out in footnotes. It is unlikely, however, that we are aware of all the terminological conventions that exist, and we apologize to the reader for any inconvenience that may result from our terminological choices.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
ABLATIVE (SPATIAL , CASE)	‘(away) from’; also ‘from above/below/inside’; marker introducing a spatial participant; direction from
ABLATIVE (TEMPORAL , CASE)	‘from’, ‘since’; marker introducing a temporal (source) participant
ACCORDING TO (CASE , CONJUNCTION)	‘according to’; marker introducing a nominal or clausal participant
ACROSS (SPATIAL , CASE)	‘across’; marker introducing a locative participant
ADDITIVE	‘plus’, ‘and’; marker introducing a quantifying participant
ADVERSATIVE ² (CONJUNCTION)	‘but’, ‘however’, ‘nevertheless’; marker introducing an adversative participant
AFTER (TEMPORAL)	‘later than’, ‘after’; marker introducing a temporal participant

² Beth Levin (personal communication) points out that there are alternative uses of the term “adversative.”

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
AGENT (CASE)	e.g., 'by'; marker for a participant that instigates or performs the action described by the main verb
AGREEMENT	marker of grammatical agreement, i.e., of the person, number, gender, or class, typically on the verb
ALLATIVE (SPATIAL)	'to'; marker introducing an allative/directional participant; direction toward
ALREADY	'already'; focus particle or marker
ALSO	'also', 'too', 'as well'; marker modifying nouns and other categories
NP-AND (CONJUNCTION)	'and'; noun phrase-conjoining marker
S-AND (CONJUNCTION)	'and'; clause-conjoining marker
ANDATIVE	'motion thither'; marker for a movement away from the speaker or deictic center; itive. Cf. VENITIVE
Anterior	see PERFECT
Antibenefactive	see MALEFACTIVE
ANTICAUSATIVE³	marker that typically reduces the valence of a verb by one participant, which as a rule is the agent
AROUND (SPATIAL, CASE)	'round about', 'round and round'; marker introducing a locative participant
AVERTIVE (ASPECT)	'almost, nearly'; marker for an action or event that was on the verge of taking place but did not take place. Cf. PROXIMATIVE
BEFORE (TEMPORAL, CASE)	'before', 'earlier'; marker introducing a temporal participant
BEHIND (SPATIAL)	'behind', 'back', 'in back of', 'after'; marker introducing a locative participant; "backterior"
BENEFACTIVE (CASE)	'for', 'for the benefit of'; marker introducing a participant indicating that the action of the main verb is for the benefit or on behalf of someone else. Cf. MALEFACTIVE

³ According to Haspelmath (1990: 33), an anticausative "denotes a spontaneous process without an implied agent, while the basic verb denotes a transitive action." Anticausative markers, which are not infrequently referred to as intransitivizing elements or intransitivizers, differ from passives in that no agent is implied.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
BESIDE (SPATIAL, CASE)	‘beside’, ‘at the side of’; marker introducing a locative participant
CASE	marker used for introducing a nominal (or pronominal) participant
CAUSATIVE	‘cause to be’, ‘cause to do’; a marker for an agent that brings about the action or state it describes
CAUSE (CASE, CONJUNCTION)	‘because of’, ‘since’, ‘on account of’, ‘therefore’; marker introducing a participant of cause or reason
CERTAINTY (EPISTEMIC MODALITY)	‘it is certain that’; marker used by the speaker to emphasize that the proposition is true
CESSATIVE (ASPECT)	indicates that an event stops but not necessarily that it is completed. Cf. COMPLETIVE
CHANGE-OF-STATE	‘become’, ‘turn into’; inchoative, ingressive. Cf. RESULTATIVE
CLASSIFIER	classificatory particle; a general term referring to the specific system of formatives that consists of quantifiers, repeaters, and noun classifiers proper (cf. Senft 1996: 16)
COMITATIVE (CASE)	‘(together) with’; marker introducing a comitative participant
COMMON (GENDER)	gender category that includes feminine and masculine, possibly also other concepts. Cf. NEUTER
COMPARATIVE (CASE)	‘than’; marker of standard in comparative constructions of inequality. See also EQUATIVE COMPARATIVE
COMPLEMENTIZER (CONJUNCTION)	‘that’; marker introducing complement clauses
COMPLETIVE (ASPECT)	indicates that something is done thoroughly and to completion. Cf. CESSATIVE
CONCERN (CASE, CONJUNCTION)	‘about’, ‘concerning’; marker introducing a nominal or clausal participant
CONCESSIVE (CONJUNCTION)	‘despite the fact that’, ‘even though’; marker introducing a concessive participant
CONDITIONAL (CONJUNCTION)	‘if’; marker of conditional protasis
CONJUNCTION	e.g., ‘and’, ‘accordingly’, ‘but’, etc.; marker used for conjoining clauses; clause connective, sentence connective

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
CONSECUTIVE (CONJUNCTION)	'and then', 'thereafter'; narrative discourse marker
CONTINUOUS (ASPECT)	'be doing', 'keep on doing'; marker for an event that is in progress at reference time; this term combines the notions of both progressive and durative aspects
COPULA ⁴	'be'; predicate marker used in propositions of the type 'X is (a) Y'; identifying copula, classifying copula. See also EXIST; LOCATIVE COPULA
DATIVE (CASE)	'to'; marker for – typically – a human recipient; indirect object
DEFINITE	'the'; definite article; nominal determiner
DEMONSTRATIVE	'this/these', 'that/those'; nominal determiner
DEONTIC (MODALITY) ⁵	is concerned with necessity or possibility of acts performed by morally responsible agents; see OBLIGATION; PERMISSIVE
DIMINUTIVE	'smaller than normal'
DISTAL (SPATIAL)	'far away'; deictic marker for spatial distance. Cf. PROXIMAL
DOWN (SPATIAL)	'down', 'below', 'under', 'underneath'; marker used to introduce a locative participant
DUAL (NUMBER)	marker for a number unit consisting of no more and no less than two items
Durative	see CONTINUOUS
EARLIER (TEMPORAL)	'earlier', 'before', 'ago'; temporal marker
EGRESSIVE (ASPECT)	'stop doing'; see also CESSATIVE
ELATIVE ⁶	'too', as in <i>too much</i> , <i>too big</i> , etc. Cf. SUPERLATIVE
EMPHATIC	marker expressing emphasis or contrast
Emphatic reflexive	see INTENSIVE-REFL
EPISTEMIC (MODALITY)	is concerned with the speaker's knowledge and beliefs about the state of affairs expressed in the utterance; see CERTAINTY; POSSIBILITY; PROBABILITY

⁴ With the term COPULA, we are referring to a range of different predicative notions, including identification, classification, specification, and characterization (see Hengeveld 1992). Excluded are existential copulas (see EXIST) and locative copulas (see LOCATIVE COPULA).

⁵ Deontic modality has also been called "agent-oriented modality" (see, e.g., Bybee et al. 1994) or "root modality" (Coates 1995).

⁶ Note that this term is used in quite a different sense in the literature on case marking, where it refers to the notion 'out of'.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
EQUATIVE	‘as . . . as’; comparative marker of equality;
COMPARATIVE	comparison of equality
EQUATIVE COPULA	‘be’, as in <i>John is a teacher</i> ; predicate marker
ERGATIVE	marker introducing the agent argument of a transitive verb in ergative languages
EVEN	‘even’; scalar focus particle
EVIDENTIAL	marker used by the speaker to indicate the source of the information on which a given assertion is based. The term is generally used to describe devices indicating perceptual evidence (both direct and indirect) and devices indicating evidence that is obtained from someone else.
EVIDENTIAL, INFERENTIAL	marker adding the following nuance of meaning to a given utterance: ‘I have evidence that it happened, and I infer that it must have happened.’
EXCLAMATION	e.g., ‘hi there!’
EXCLUSIVE	‘we excluding you’; a distinction made within (>) FIRST PERS-PRON, which excludes the hearer/addressee. Cf. INCLUSIVE
EXIST⁷	‘there is [X]’, ‘[X] exists’
FEMALE	‘female’; marker used as a nominal modifier to refer to female participants
FIRST (NUMERAL)	‘(the) first’; ordinal numeral
FIRST (PERS-PRON)	‘I’, ‘we’; first person pronoun
FIRST (TEMPORAL)	‘at first’, ‘to begin with’
FOCUS	marker used in sentences that focus on some participant, typically presenting that participant as new information
FREQUENTATIVE (ASPECT)	marker indicating that an event takes place frequently, i.e., neither once nor habitually
FRONT (SPATIAL, CASE)	‘in front of’, ‘before’; marker introducing a locative participant; “fronterior”
FUTURE (TENSE)	‘will’, ‘shall’; indicates that the speaker predicts an event to occur after the moment of speech

⁷ EXIST includes what Hengeveld (1992) refers to as existence and reality. EXIST markers are typically one-argument predicates (e.g., *There is coffee*); however, they can also have two participants (e.g., roughly, *There is coffee for you*), which differ drastically from one-participant markers in their grammaticalization behavior.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
FUTURE, NEAR (TENSE)	indicates that the speaker predicts an event to occur very soon after the moment of speech; near future, immediate future
HABITUAL (ASPECT)	'do habitually'; marker for an event occurring habitually or usually, repeated on different occasions
HONORIFIC	marker of honorific reference
HORTATIVE	marker used by the speaker to encourage or incite someone to action
IMMEDIATE	see FUTURE, NEAR; PAST, NEAR
IMPERFECTIVE (ASPECT)	marker used to indicate that an event is viewed as unbounded temporally. Cf. PERFECTIVE
IMPERSONAL	marker for an agent that is suppressed but still understood
IN (SPATIAL)	'in', 'inside', 'within'; marker introducing a locative participant; interior
IN (TEMPORAL)	'in', 'within', 'during'; marker introducing a temporal participant
INCEPTIVE (ASPECT)	'start doing', 'begin doing'; inceptive, ingressive
Inchoative	see CHANGE-OF-STATE
INCLUSIVE	'we including you'; a distinction made within (>) FIRST PERS-PRON, which includes the hearer/addressee; cf. EXCLUSIVE
INDEFINITE	'a, an'; indefinite article; nominal determiner
INDEFINITE PRONOUN	'something', 'someone', etc.
Ingressive	see CHANGE-OF-STATE
INSTEAD (CASE, CONJUNCTION)	'instead of'; marker introducing a nominal or clausal participant; replative
INSTRUMENT (CASE)	'with', 'by means of'; marker used to present a participant as an instrument
INTENSIFIER	'very', 'extremely'
INTENSIVE-REFL	'-self', as in <i>The king himself</i> , <i>The king did it himself</i> ; emphatic reflexive, intensifier, identifier
INTENTION	'to intend to'
Interrogative	see S-QUESTION, W-QUESTION
ITERATIVE (ASPECT)	'do repeatedly'; repetitive; marker indicating that an action is repeated
LATE (TEMPORAL)	'be late (be delayed)'
LATER (TEMPORAL)	'then', 'thereafter', 'afterwards', 'later'
LOCATIVE	marker introducing a locative participant

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
LOCATIVE COPULA	'be at', 'be somewhere'; predicate marker used in propositions of the type 'X is (located) at Y'
LOGOPHORIC	marker used in indirect quotes referring to the person being quoted; designating a particular category of anaphoric pronouns, personal and possessive, which refer to the author of a discourse or to a participant whose thoughts are reported
MALE	'male'; marker used as a nominal modifier to refer to male participants
MALEFACTIVE (CASE)	'to the detriment of'; marker for a participant indicating that the action of the main verb is to the detriment of someone else; antibenefactive. Cf. BENEFACTIVE
MANNER (CASE, CONJUNCTION)	marker introducing a manner participant
MATERIAL (CASE)	'from', 'with'; marker for a participant typically indicating the material from which an object is made
MIDDLE ⁸	marker indicating that the patient of the action is implicated as contributing to the action in some way
MIRATIVE ⁹	marker used for utterances reporting information that is new or surprising to the speaker regardless of whether the information source is first- or secondhand
NEGATION	'not', 'no'; marker of negation
NEGATION, EXIST	'there is not/no'
NEUTER (GENDER)	a gender category that is neither feminine nor masculine. Cf. COMMON
NEXT	'the next', 'the following'
NO	'no'; interjection

⁸ Kemmer (1993: 238) observes, "The semantic middle is a coherent but relatively diffuse category that comprises a set of loosely linked semantic sub-domains centering roughly around the direct reflexive." It remains unclear whether we are really dealing with a distinct functional notion (Beth Levin, personal communication); we are including it tentatively on account of the discussion in Kemmer 1993.

⁹ Here we accept the standpoint taken by DeLancey 1997 that the mirative represents a category of its own. This view is radically different from the one presented in Lazard 1999, where the mirative is treated as one of the three "values" of a more abstract category of "mediative," the other two values being hearsay and inference.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
NO LONGER	'no longer'
NOT YET	'not yet'
NP-and	<i>see</i> AND
Object marker	<i>see</i> PATIENT
OBLIGATION (DEONTIC MODALITY)	'have to', 'should', 'must'; the agent is presented as being obliged to perform the action of the main verb
OBVIATIVE	marker indexing a change in the subject; switch reference
ONE (NUMERAL)	'one'; cardinal numeral
ONLY	'alone', 'merely', 'just'
OPTATIVE	the proposition represents the speaker's will
OR (CONJUNCTION)	'or'; alternative marker, conjoining noun phrases or clauses
OTHER	'another', 'other'
OUT (SPATIAL)	'out', 'outside'
PARTITIVE (CASE)	marker introducing a participant expressing the notion 'a part of' or 'partly affected'
PASSIVE	a marker indicating that the action is viewed from the perspective of the recipient or patient of the verb, while the agent is suppressed or demoted
PAST (TENSE)	indicates that an event occurs before the moment of speech
PAST, NEAR (TENSE)	an event that occurred immediately before the moment of speech; recent past, near past, immediate past
PATH (SPATIAL, CASE)	'through', 'via'; marker introducing a locative participant; path marker
PATIENT (CASE)	marker for a participant that is the undergoer of the action denoted by the verb; direct object
PERFECT¹⁰ (ASPECT)	marker indicating that a past event is relevant to the situation at reference time; anterior
PERFECTIVE (ASPECT)	marker used to indicate that an event is viewed as bounded temporally. Cf. IMPERFECTIVE
PERMISSIVE (DEONTIC MODALITY)	'be allowed to'; the agent is allowed to do the action of the main verb

¹⁰ Our term "perfect" corresponds to what Bybee et al. (1994) call the "anterior."

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
PERS-PRON (PRONOUN)	personal pronoun, pronominal marker. See also FIRST ; SECOND ; THIRD
PLURAL (NUMBER)	plural marker, typically on nouns
A-POSSESSIVE	'of'; marker of attributive (nominal) possession; genitive case, associative, connective, nominal possessive. (For description of term, see Heine 1997a.)
B-POSSESSIVE	'X belongs to Y', 'X is Y's'; predicative possession, marker of <i>belong</i> -constructions. (For description of term, see Heine 1997a.)
H-POSSESSIVE	'have', 'own'; predicative possession, marker of possessive <i>have</i> -constructions. (For description of term, see Heine 1997a.)
POSSIBILITY (EPISTEMIC MODALITY)	'it is possible that'; marker expressing that the speaker indicates that the situation described in the proposition is possibly true
PRESENT (TENSE)	marker indicating an event is occurring simultaneously with the moment of speech
PROBABILITY (EPISTEMIC MODALITY)	'it is likely that'; with such markers, the speaker indicates that the situation described in the proposition is probably true
Progressive	see CONTINUOUS
PROHIBITIVE	'don't do!'; negative imperative
PRONOUN	a marker standing for a noun or noun phrase
PRO-VERB	semantically empty predicate marker standing for other verbs in certain contexts; e.g., <i>do</i> as in <i>do jogging</i>
PROXIMAL (SPATIAL)	'nearby', 'close to'; deictic marker for spatial proximity. Cf. DISTAL
PROXIMATIVE (ASPECT)¹¹	'be about to', i.e., 'be on the verge of doing'. Cf. AVERTIVE
PURPOSE (CASE, CONJUNCTION)	'in order to', 'so that'; a marker introducing the purpose of an action
S-QUESTION	marker of polar (yes-no) questions
W-QUESTION	'who?', 'what?', etc.; marker of word questions
QUOTATIVE	a marker introducing direct speech
RECIPROCAL (PRONOUN)	'each other'; a marker indicating that participants act upon each other

¹¹ Note that this term is also used in some other ways; here it refers exclusively to an aspectual notion (see Heine 1994b).

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
REFLEXIVE (PRONOUN)	'self', as in <i>I saw myself in the mirror</i> ; the patient is the same entity as the agent (i.e., the two have identical reference)
RELATIVE (CONJUNCTION)	'who', 'which', 'that'; marker introducing relative clauses
Repetitive	see ITERATIVE
RESULTATIVE (ASPECT)¹²	'having reached a new state'. Cf. CHANGE-OF-STATE
S-and	see AND
S-question	see QUESTION
SAME	'(the) same', 'identical'
SECOND (PERS-PRON)	'you', 'you all'; second person pronoun
SIDE (CASE)	'by the side of', 'on the side of'; marker introducing a locative participant
SIMILE (CASE, CONJUNCTION)	'like', 'as if', 'thus'; marker of simile or similarity participants; similitive
SINCE (TEMPORAL, CASE, CONJUNCTION)	'since (the time when)'; marker introducing temporal participants
SINCE (CAUSAL, CONJUNCTION)	'since, as, because'; marker introducing a causal participant
SINGULATIVE (NUMBER)	marker restricting the reference (of a noun) to a single entity
SOME (QUANTIFIER)	'some'; approximative marker
SPATIAL (CASE)	marker introducing a spatial/locative participant
STILL	'still'; focus particle or marker
SUBORDINATOR (CONJUNCTION)	marker introducing adverbial clauses
SUCCEED¹³	'manage to do', 'succeed in doing'
SUPERLATIVE	'(the) most'; marker for 'a position on top of or over'. Cf. ELATIVE
TEMPORAL	marker introducing a temporal participant
Terminative	see EGRESSIVE
THEN (TEMPORAL)	'then', 'afterwards', 'later'
THERE (SPATIAL)	'there'; deictic marker of distal location. Cf. DISTAL
THIRD (PERS-PRON)	'he', 'she', 'it', 'they'; third person pronoun

¹² **RESULTATIVE** is also used in other senses; here we use it exclusively as a term for a verbal aspect. Conceivably, **RESULTATIVE** and (>) **CHANGE-OF-STATE** can be grouped together.

¹³ While 'succeed' is typically encoded as a lexical item, some languages appear to treat it as a functional category.

<i>Concept Label</i>	<i>Approximate Gloss and Descriptive Notes</i>
TOGETHER	'together'
TRANSITIVIZER	marker transforming an intransitive verb into a transitive one
TRIAL (NUMBER)	marker for a number unit consisting of no more and no less than three items
TWO (NUMERAL)	'two'; cardinal numeral
UNTIL (TEMPORAL, CASE, CONJUNCTION)	'until', 'up to'; marker introducing a temporal participant
UP (SPATIAL)	'up', 'on', 'above', 'over'; marker introducing a locative participant; "superior"
VENITIVE	'motion hither', 'motion towards'; marker for a movement toward the speaker or deictic center; ventive. Cf. ANDATIVE
VP-and	see AND
W-question	see QUESTION

Source–Target Lexicon

A

‘Abandon’ see LEAVE¹

ABILITY > (1) PERMISSIVE

This is a well-researched instance of grammaticalization (see, e.g., Traugott 1972: 198–9; Kytö 1987; Bybee et al. 1991: 25; Bybee et al. 1994: 187–94; Table 6.3). Old Chinese (*de* ‘to obtain’ >) *de* ability marker > permissive marker. Ex.

Middle Chinese (tenth century A.D.; Zutangji 5/98/7; quoted from Sun 1996: 121)

- | | | | | | | |
|-----|------------|------------|------------|-----------|----------------|------------|
| (a) | <i>hai</i> | <i>jie</i> | <i>pan</i> | <i>de</i> | <i>xu-kong</i> | <i>bu?</i> |
| | still | explain | judge | possible | empty | NEG |
- ‘Can (you) still tell what emptiness is?’

Middle Chinese (tenth century A.D.; Zutangji 1/153/3; quoted from Sun 1996: 124)

- | | | | | | |
|-----|-----------|-----------|-----------|------------|------------|
| (b) | <i>ni</i> | <i>de</i> | <i>ru</i> | <i>men</i> | <i>ye.</i> |
| | you | possible | enter | door | PART |
- ‘You may enter the door (to join).’

Archaic Chinese *neng* ‘be able’, ‘be capable’ > marker of possibility and permission (Alain Peyraube, personal communication). English *may* have started out with a meaning of physical ability or power and has come to be used to report permission (Bybee et al. 1994: 193). German *können* ‘to be able’ > ‘to be allowed to’. Ex.

German

- | | | | | |
|-----|------------|-------------|-------------|------------------|
| (a) | <i>Ich</i> | <i>kann</i> | <i>Auto</i> | <i>fahr- en.</i> |
| | I | can | car | drive-INF |
- ‘I know how to drive.’

¹ Concerning the meaning of grammatical concepts, see the list of grammatical concepts in Chapter 2.