CAMBRIDGE TEXTBOOKS IN LINGUISTICS

Grammaticalization

Second Edition

Paul J. Hopper and
Elizabeth Closs Traugott

Grammaticalization

Second Edition

This is a general introduction to grammaticalization, the change whereby lexical terms and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions. Thus nouns and verbs may change over time into grammatical elements such as case markers, sentence connectives, and auxiliaries. The authors synthesize work from several areas of linguistics, including historical linguistics, discourse analysis, and pragmatics. Data are drawn from many languages including Ewe, Finnish, French, Hindi, Hittite, Japanese, Malay, and especially English.

This second edition has been thoroughly revised with substantial updates on theoretical and methodological issues that have arisen in the decade since the first edition, and includes a significantly expanded bibliography. Particular attention is paid to recent debates over directionality in change and the role of grammaticalization in creolization.

Grammaticalization will be a valuable and stimulating textbook for all linguists interested in the development of grammatical forms. Readers in anthropology and psychology will also appreciate the insights it offers into the interaction of language and structure and use.

PAUL J. HOPPER is Paul Mellon Distinguished Professor of Humanities at Carnegie Mellon University. His publications include *Grammaticalization* (co-authored with Elizabeth Closs Traugott, Cambridge, 1993), *A Short Course in Grammar* (1999), *The Limits of Grammaticalization* (co-edited with Anna Giacalone-Ramat, 1998), and *Frequency and the Emergence of Linguistic Structure* (co-edited with Joan Bybee, 2001).

ELIZABETH CLOSS TRAUGOTT is Professor of Linguistics and English at Stanford University. Her publications include *A History of English Syntax* (1972), *Linguistics for Students of Literature* (co-authored with Mary L. Pratt, 1980), *Grammaticalization* (co-authored with Paul J. Hopper, Cambridge, 1993), and *Regularity in Semantic Change* (co-authored with Richard B. Dasher, Cambridge, 2001).

CAMBRIDGE TEXTBOOKS IN LINGUISTICS

General editors: P. AUSTIN, J. BRESNAN, B. COMRIE,
W. DRESSLER, C. J. EWEN, R. LASS, D. LIGHTFOOT,
I. ROBERTS, S. ROMAINE, N. V. SMITH

In this series

P. H. MATTHEWS *Morphology* Second edition
B. COMRIE *Aspect*
R. M. KEMPSON *Semantic Theory*
T. BYNON *Historical Linguistics*
J. ALLWOOD, L.-G. ANDERSON and Ö. DAHL *Logic in Linguistics*
D. B. FRY *The Physics of Speech*
R. A. HUDSON *Sociolinguistics* Second edition
A. J. ELLIOTT *Child Language*
P. H. MATTHEWS *Syntax*
A. RADFORD *Transformational Syntax*
L. BAUER *English Word-Formation*
S. C. LEVINSON *Pragmatics*
G. BROWN and G. YULE *Discourse Analysis*
R. HUDDLESTON *Introduction to the Grammar of English*
R. LASS *Phonology*
B. COMRIE *Tense*
W. KLEIN *Second Language Acquisition*
A. J. WOODS, P. FLETCHER and A. HUGHES *Statistics in Language Studies*
D. A. CRUSE *Lexical Semantics*
A. RADFORD *Transformational Grammar*
M. GARMAN *Psycholinguistics*
G. G. CORBETT *Gender*
H. J. GIEGERICH *English Phonology*
R. CANN *Formal Semantics*
J. LAVER *Principles of Phonetics*
F. R. PALMER *Grammatical Roles and Relations*
M. A. JONES *Foundations of French Syntax*
A. RADFORD *Syntactic Theory and the Structure of English: a Minimalist Approach*
R. D. VAN VALIN, JR, and R. J. LAPOLLA *Syntax: Structure, Meaning and Function*
A. DURANTI *Linguistic Anthropology*
A. CRUTTENDEN *Intonation* Second edition
J. K. CHAMBERS and P. TRUDGILL *Dialectology* Second edition
C. LYONS *Definiteness*
R. KAGER *Optimality Theory*
J. A. HOLM *An Introduction to Pidgins and Creoles*
C. G. CORBETT *Number*
C. J. EWEN and H. VAN DER HULST *The Phonological Structure of Words*
F. R. PALMER *Mood and Modality* Second edition
B. J. BLAKE *Case* Second edition
E. GUSSMAN *Phonology*
M. YIP *Tone*
W. CROFT *Typology and Universals* Second edition
F. COULMAS *Writing Systems: an introduction to their linguistic analysis*
L. WHITE *Second Language Acquisition and Universal Grammar*
P. J. HOPPER and E. C. TRAUGOTT *Grammaticalization* Second edition

Grammaticalization

Second Edition

PAUL J. HOPPER

Paul Mellon Distinguished Professor of Humanities,
Carnegie Mellon University

ELIZABETH CLOSS TRAUGOTT

Professor of Linguistics and English,
Stanford University

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, 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

© Paul J. Hopper, Elizabeth Closs Traugott 1993, 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1993
Second edition 2003

Typeface Times and Formata Regular 10/13 pt. *System* LaTeX 2_ε [TB]

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

Library of Congress Cataloguing in Publication data applied for

ISBN 0 521 00948 0 hardback
ISBN 0 521 80421 3 paperback

Transferred to digital printing 2004

In memory of
Dwight Bolinger
1907–1992

Contents

List of figures page xiii
List of tables xiv
Preface and acknowledgments xv
List of abbreviations xviii

1 Some preliminaries 1
1.1 Introduction 1
1.2 What is a grammaticalized form? 4
1.2.1 A preliminary classification of grammatical forms 4
1.2.2 Clines 6
1.2.3 Periphrasis versus affixation 7
1.3 Some further examples of grammaticalization 9
1.3.1 *Lets* 10
1.3.2 A West African complementizer 13
1.3.3 Agreement markers 15
1.4 Grammaticalization and language structure 16
1.5 Grammaticalization and the directionality of language change 16
1.6 Conclusion 17

2 The history of grammaticalization 19
2.1 Introduction 19
2.2 Earlier research on grammaticalization 19
2.3 Research on grammaticalization from the 1960s to the 1990s 25
2.4 Recent trends in research on grammaticalization 30

3 Mechanisms: reanalysis and analogy 39
3.1 Introduction 39
3.2 Some background assumptions about change 40
3.2.1 Induction, deduction, abduction 41
3.2.2 Who is the language learner? 43
3.2.3 The question of genetic endowment 45
3.2.4 Innovation versus spread 46

3.3 Reanalysis 50
3.3.1 The French inflectional future 52
3.3.2 The English modal auxiliaries 55
3.4 The independence of reanalysis and grammaticalization 58
3.4.1 Word-order change 59
3.5 Analogy/rule generalization 63
3.6 The differential effects of reanalysis and analogy 68
3.7 Conclusion 69

4 Pragmatic factors 71
4.1 Introduction 71
4.2 Inferencing and meaning change 74
4.2.1 Semantics and pragmatics 76
4.2.2 Relationships between senses of a form: homonymy and polysemy 77
4.2.3 Conversational and conventional inferencing 78
4.3 The role of pragmatic inferencing in grammaticalization 81
4.3.1 Metaphorical processes 84
4.3.2 Metonymic processes 87
4.4 Metaphor and metonymy as problem solving 92
4.5 Pragmatic enrichment versus "bleaching" 94
4.6 Conclusion 98

5 The hypothesis of unidirectionality 99
5.1 Introduction 99
5.2 Generalization 100
5.2.1 Generalization of meaning 101
5.2.2 Generalization of grammatical function 104
5.3 Decategorialization 106
5.3.1 A noun-to-affix cline 110
5.3.2 A verb-to-affix cline 111
5.3.3 Multiple paths 114
5.4 Some processes participating in unidirectionality 115
5.4.1 Specialization 116
5.4.2 Divergence 118
5.4.3 Renewal 122
5.5 A synchronic result of unidirectionality: layering 124
5.6 Frequency
5.6.1 Frequency effects 127
5.6.2 Synchronic studies of frequency 128
5.6.3 Diachronic studies of frequency 129
5.7 Counterexamples to unidirectionality 130
5.8 The uses of unidirectionality in reconstruction 138
5.9 Conclusion 139

6 Clause-internal morphological changes 140
6.1 Introduction 140
6.2 Morphologization 140
6.2.1 Some characteristics of clitics 142
6.2.2 Positions of clitics 143
6.2.3 Semantic "relevance" as a factor in fusion and morpheme order 151
6.2.4 Phonological concomitants of morphologization 154
6.3 The development of paradigms 159
6.4 Argument-structure marking: functional-semantic hierarchies and morphological generalization 165
6.4.1 Object marking in Persian 165
6.4.2 Ergative case marking: a statistical perspective 168
6.5 Loss 172
6.6 Conclusion 174

7 Grammaticalization across clauses 175
7.1 Introduction 175
7.2 A cline of clause-combining constructions 176
7.2.1 Parataxis 179
7.2.2 Hypotaxis 181
7.2.3 Subordination 183
7.3 The grammaticalization of clause linkers 184
7.4 Examples of the development of complex sentence constructions 190
7.4.1 Tftaf-complementation in English 190
7.4.2 Quotative say-constructions in Akkadian 194
7.4.3 Relative clauses in English and Hittite 196
7.5 From complex to simple clauses 204
7.5.1 From clause chaining to verb inflection in Lhasa 204
7.5.2 Two conjoined clauses reanalyzed as a single clause 206
7.5.3 From main clause construction to sentential adverb in contemporary English 207
7.6 Some counterexamples to unidirectionality in clause combining 209
7.7 Conclusion 211

8 Grammaticalization in situations of extreme language contact 212
8.1 Introduction 212
8.2 Basic characteristics of pidgins and creoles 213
8.2.1 Some characteristics of pidgins 214
8.2.2 Some characteristics of creoles 216

8.3 Implications of pidgins and creoles for language change 219
8.3.1 Child versus adult language acquisition 219
8.3.2 Simplification and elaboration 222
8.4 Specific implications of pidgins and creoles for grammaticalization 224

9 Summary and suggestions for further work 231

Notes **234**
References **237**
Index of names **265**
Index of languages **270**
General index **272**

Figures

3.1 A model of language change *page* **41**
3.2 Schema of the development of auxiliary be going to **69**
4.1 Revised schema of the development of auxiliary *be going to* **93**
7.1 Properties relevant to the cline of clause combining **179**

Tables

3.1 **Grammaticalization of VO word order in English between AD 1000 and AD 1500** *page* **67**
5.1 **Approximate proportion of compound verbs in Indo-Aryan languages** **113**
5.2 **Ratio of compound verbs in Marathi and Hindi-Urdu according to semantic class of main verb** **114**
6.1 **Buryat Mongolian pronouns and verb endings** **141**
6.2 **Polish tonic and clitic forms of the copula** **146**
6.3 **Bonding of clitic copula to verb stem in Polish, AD 1500 to the present** **146**
6.4 **Differential univerbation of preterit verb and person-number suffix in Polish dialects** **147**
6.5 **Affixal aspect-tense-mood forms** **154**
6.6 **Old Icelandic present indicative reflexive verb forms** **160**
6.7 **Pre-Sanskrit noun inflection** **162**
6.8 **Sanskrit and pre-Pali forms of the copula** **163**
6.9 **Pre-Pali and Pali forms of the copula** **163**
6.10 **Differential inflection of the aorist in Bulgarian dialects** **164**
6.11 **Differential inflection of past participles in Bulgarian dialects** **164**
6.12 **Old English strong adjective singular inflection** **172**
7.1 **Occurrence of *that* with *think* and *guess* versus all other verbs** **209**

Preface to the second edition and acknowledgments

When we wrote the first edition of this book in the early 1990s, our aim was to present an overview of grammaticalization for the benefit of those students of linguistics to whom this was a new or only vaguely familiar framework for understanding linguistic phenomena. We defined grammaticalization as the process whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalized, continue to develop new grammatical functions. We also characterized it as the process whereby the properties that distinguish sentences from vocabulary come into being diachronically or are organized synchronically. In the last ten years grammaticalization has become a major field of study, and work on the topic has flourished in both "functional" and "formal" frameworks, that is, on the one hand frameworks that privilege the interplay of language structure and use and consider language as both a cognitive and a communicative force, and on the other frameworks that privilege language structure and consider language primarily from the perspective of internalized systems. As work has progressed, it has become clear that the definition of grammaticalization as a "process" has been misleading. To some it has suggested that grammaticalization is conceived as a force with an impetus of its own independent of language learners and language users. This was never intended. Only people can change language. The challenge presented by the kinds of data of central concern to those who work on grammaticalization is that morphosyntactic changes are replicable across languages and across times; furthermore, they have a very strong tendency to go in the same direction, e.g. lexical verb to auxiliary, nominal adpositions to case marker, not vice versa. To avoid further terminological confusion, we now define grammaticalization as the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions.

In this second edition we have maintained the basic structure and content of the original book, but because grammaticalization has come to be so widely studied in the last decade we have added substantial discussion of recent issues where they seemed germane. Linguistic theory has also undergone some radical changes since the book was originally written; for example, within generative theory rule-based

grammars have been challenged by constraints-based grammars (see developments in Optimality Theory sparked by A. Prince and Smolensky 1997); and issues of the relationship between structure and use have come to be of central concern to most theoretical perspectives on linguistics. We have not attempted to address the variety of issues that form the current theoretical dialog. That would require not only revision, but reconceptualization, of the original book. Instead, we have maintained the main outlines of the original work, deleted some obviously outdated analyses and suggested pointers where relevant to newer ones.

Many terms in linguistics serve two functions, one to describe properties of language (e.g., syntax, morphology, phonology), the other to name the study and theory of that property. The term grammaticalization is no exception. It refers not only to changes observable across time, but also to an approach to language study, one that highlights the interaction of use with structure, and the non-discreteness of many properties of language. In this book we do not advance any particular theory of language, although the framework can be characterized as "integrative functionalist" (Croft 1995) in that we consider linguistic phenomena to be systematic and partly arbitrary, but so closely tied to cognitive and social factors as not to be self-contained. We focus on observations that lead to claims about what sorts of concepts an adequate theory must account for. We discuss the kinds of interpretation of data that flow from the approach of grammaticalization, as well as the kinds of data that one would look for in studying some aspect of a language from this perspective. The study of grammaticalization touches on many of the topics that have been central to work in linguistics, whether synchronic or diachronic, most particularly the domains of morphosyntax and morphology. No attempt is made in this book to cover every topic in an encyclopedic way, nor could it be. Some basic knowledge of linguistics is needed. We assume that readers have at least worked through one of the standard introductory textbooks, and have either had a course in historical linguistics or have carefully read a recent textbook in this field.

Our deepest thanks to the numerous people who contributed directly or indirectly to either the first or the second edition of this book, or both. Among them Luc Baronian, Joan Bybee, Brady Clark, William Croft, Andrew Garrett, A. Gianto, T. Givón, the late Joseph Greenberg, Claude Hagège, Bernd Heine, Suzanne Kemmer, Paul Kiparsky, Christian Lehmann, Douglas Lightfoot, Therese Lindström, John McWhorter, Fritz Newmeyer, John Rickford, Sarah Roberts, Devyani Sharma, Scott Schwenter, Eve Sweetser, Sandra Thompson, and Max Wheeler deserve special mention. Olga Fischer and her students, especially Marije Bogers, Anthony Glass, and Brad Philpot, drew our attention to many points in the first edition that deserved clarification, as did Claire Cowie. We have been especially indebted to the late Suzanne Fleischman and to Nigel Vincent, who commented on drafts of the first edition, and to Roger Lass and Arnold Zwicky,

who made detailed comments on the published version. Their careful and insightful comments were of inestimable value. Thanks too to the many students and faculty who attended our seminar at the Linguistic Society of America's Linguistic Institute at Stanford University in 1987, where many of the ideas for the writing of this book were developed, and in subsequent courses that we have taught in our respective institutions and at summer sessions around the world. Julia Harding and John McWhorter provided invaluable help with editorial matters for the first edition, Kay McKechnie, Rob Podesva, and Elyse Nakajima for the second. We would also like to thank Penny Carter and Judith Ayling for their help in getting the first edition to press, and Andrew Winnard for his support of the second edition.

We acknowledge the opportunities provided by Guggenheim Fellowships (Hopper in 1985, Traugott in 1983) for the study of many of the aspects of grammaticalization discussed here. Elizabeth Traugott further acknowledges a Fellowship at the Center for Advanced Study of the Behavioral Sciences at Stanford (1983).

In writing this book we debated many points with respect for each other's different views. No two authors covering as wide a range of issues currently open to theoretical and methodological debate as are touched on in this book could expect to agree with every concept put forward, and we are no exception. We have welcomed the inspiration that our differing perspectives have provided, and we hope that this book will encourage further debate in the linguistic community.

Abbreviations

Linguistic terms

ABL	ablative
ABS	absolutive
ACC	accusative
ADV	adverb
AGR	agreement marker
AGT	agent
AOR	aorist
ASP	aspect
AUX	auxiliary
BEN	benefactive
CAUS	causative
CL	noun classifier
COMP	complementizer
COMPL	completive
CONJ	conjunction
CONTIN	continuative
COP	copula
DAT	dative
DEF	definite marker
DEM	demonstrative
DET	determiner
DUR	durative
EMPH	emphatic
ERG	ergative
FEM	feminine
FUT	future
GEN	genitive
GER	gerundive

ILL	illative
IMP	imperative
IMPF	imperfect
INCL	inclusive
INDEF	indefinite
INF	infinitive
INSTR	instrumental
IOBJ	indirect object
LOC	locative
M	modal
MASC	masculine
MED	medial
NEG	negative
NEUT	neuter
NF	non-final marker
NOM	nominative
NP	noun phrase
NPrel	anaphoric NP in a relative clause
OBJ	object
OM	object marker
OV	object–verb word order
P	particle
PART	participle
PARTIT	partitive
PERF	perfect
PL	plural
PM	predicate marker
PN	personal name
POSS	possessive
POST	postposition
PREP	preposition
PRES	present
PRO	pronoun
PROG	progressive
PURP	purposive
QUOT	quotative marker
RECIP	reciprocal
REFL	reflexive
REL	relative
SG	singular

Srel	relative clause
SUB	subordinator
SUBJ	subject
SUBJUNCT	subjunctive
SW	switch-reference marker
TNS	tense
TOP	topic
TRANS	transitive
V	verb
VO	verb–object word order
VP	verb phrase
1	first person
2	second person
3	third person

Symbols

>	becomes/is replaced by/splits into
<	derives from
-	morpheme boundary
:	portmanteau morpheme boundary
~	alternates with

Stages of English

OE	Old English (*c.* 600–1125)
ME	Middle English (*c.* 1125–1500)
EMdE	Early Modern English (*c.* 1500–1750)
MdE	Modern English (*c.* 1750–1950)
PDE	Present-Day English (*c.* 1950–)

By convention, vowel length signs have been omitted in Latin and Old English forms.

1

Some preliminaries

1.1 Introduction

(1) Bill is going to go to college after all.

What is the relationship between the two instances of *go* in this sentence? The first *go* is usually analyzed as an auxiliary, the second as a main verb. Are they different morphemes that just happen to look and sound alike, that is, are they homonyms? Are they variants of the same morpheme in different contexts, that is, are they polysemous? Is the auxiliary historically derived from the main verb, and, if so, is this kind of derivation cross-linguistically attested?

What permits the pair in (2) but not the (b) sentence in (3)?

(2) a. Bill is going to go to college after all.
b. Bill's gonna go to college after all.

(3) a. Bill's going to college after all.
b. *Bill's gonna college after all.

These questions and many others are characteristic of the study of grammaticalization. As a first approximation, the answer is that the auxiliary which expresses immediate futurity derives historically from the motion verb *go* in a highly specific context, and that the two coexistent forms used to be polysemous. Such meaning–form correlations are found in a wide number of languages around the world.

The term "grammaticalization" has two meanings, one to do with a research framework within which to account for language phenomena, the other with the phenomena themselves. In this respect the term "grammaticalization" resembles not only other terms in linguistics such as "grammar," "syntax," and "phonology," but the terminology of all higher-level concepts in scholarly disciplines. As a term referring to a research framework, "grammaticalization" refers to that part of the study of language change that is concerned with such questions as how lexical items and constructions come in certain linguistic contexts to serve grammatical functions or how grammatical items develop new grammatical functions. This research framework is also concerned with characterizing the subset of cross-linguistically

recurring correlations across time among semantic–pragmatic, morphosyntactic, and (sometimes) phonological changes. It highlights the tension between the fixed and the less fixed in language, between relatively unconstrained lexical (semantic) structure and more constrained syntactic, morphosyntactic, and morphological structure. It provides the conceptual context for a principled account of the relative indeterminacy in language and of the basic non-discreteness of categories. As a term referring to actual phenomena of language, "grammaticalization" refers most especially to the steps whereby particular items become more grammatical through time. Grammaticalization in this sense is part of the wider linguistic phenomenon of structuration, through which combinations of forms may in time come to be fixed in certain functions.

Since Saussure, many linguists have approached language from one of two perspectives: that of its structure at a single point in time ("synchronic") and that of change between two or more points in time (historical or "diachronic"). The synchronic dimension of a language is said to be its system of grammatical units, rules, and lexical items (together with their meanings), that is, its grammar. It is usually conceived as essentially stable and homogeneous. The diachronic dimension, on the other hand, is understood as the set of changes linking a synchronic state of a language to successive states of the same language. The discreteness of categories and rules, and the rigidity of the distinction between the synchronic and diachronic dimensions have been called into question by work on the structured variation to be found in various social contexts, and analysis of discourse and language in use. They are also called into question by the study of grammaticalization.

Grammaticalization likewise has been studied from these two perspectives. The chief perspective is historical, investigating the sources of grammatical forms and the typical steps of change they undergo. From this perspective, grammaticalization is usually thought of as that subset of linguistic changes whereby a lexical item or construction in certain uses takes on grammatical characteristics, or through which a grammatical item becomes more grammatical. The other perspective is more synchronic, seeing grammaticalization as primarily a syntactic, discourse pragmatic phenomenon, to be studied from the point of view of fluid patterns of language use. In this book we will combine these two points of view, but with greater emphasis on the historical dimension.

Our example of *be going to/be gonna* illustrates several factors typical of grammaticalization viewed form the historical perspective:

(a) The change occurs only in a very local context, that of purposive directional constructions with non-finite complements, such as *I am going to marry Bill* (i.e., *I am leaving/traveling in order to marry Bill*). It does not occur in the context of

directionals in which the locative adverb is present, such as *I am going to London* or even *I am going to London to marry Bill.*

(b) The change is made possible by the fact that there is an inference of futurity from purposives: if I am traveling in order to marry, the marriage will be in the future. In the absence of an overt directional phrase, futurity can become salient.

(c) The shift from purposive *be going* (*to*...) to auxiliary *be going to* involves reanalysis not only of the *be going to* phrase but of the verb following it. Thus [I am going [to marry Bill]] is rebracketed as [I [am going to] marry Bill]. It also involves a change from progressive aspect to "immediate future."[1]

(d) The reanalysis is discoverable, that is, is manifest, only when the verb following *be going to* is incompatible with a purposive meaning, or at least unlikely in that context, for example, *I am going to like Bill, I am going to go to London.* In other words, the reanalysis is discoverable only because the contexts in which *be going to* can occur have been generalized, or analogised, to contexts that were unavailable before.

(e) Once the reanalysis has occurred, *be going to* can undergo changes typical of auxiliaries, such as phonological reduction. The reduction of the three morphemes *go-ing to* into one (*gonna*) is possible only because there is no longer a phrasal boundary between *-ing* and *to*.

(f) The various stages of grammaticalization of *be going* (*to*...) coexist in Modern English, although the change originates in the fifteenth century or perhaps even earlier.

(g) The original purposive meaning continues to constrain the use of the auxiliary: *be gonna* is the future of intention, plan, or schedule. As an original aspectual, it can occur in constructions where a future formed with *will* cannot:

(4) a. If interest rates are going to climb, we'll have to change our plans.
b. *If interest rates will climb, we'll have to change our plans.

This property of persistence of meaning presumably derives in part from the fact that the older *be going* (*to*...) for a long time was polysemous with and coexisted with the newer use, and hence allowed reinforcement of older meanings.

(h) The main verb *go* is relatively general in meaning, that is, it expresses any kind of motion away from the speaker, including walking, meandering, running, riding, etc.

(i) As grammaticalization has taken place, some of the original relatively concrete meaning of *go* has been lost, specifically motion and directionality. However, some new meanings have also been added; these are more abstract and speaker-based meanings, specifically temporal meanings based in speaker time. The historical development of the construction will be discussed more fully in Chapter 4.

1.2 What is a grammaticalized form?

As is usually the case with words rich in implications, there are a number of different conceptions of grammaticalization. Yet there are central, prototypical instances of grammaticalization which most linguists would recognise, and we start with some of them.

For example, it is usually accepted that some kind of distinction can be made in all languages between "content" words (also called "lexical items," or "contentives"), and "function" words (also called "grammatical" words). The words *example*, *accept*, and *green* (i.e., nouns, verbs, and adjectives) are examples of lexical items. Such words are used to report or describe things, actions, and qualities. The words *of*, *and*, *or*, *it*, *this*, that is, prepositions, connectives, pronouns, and demonstratives, are function words. They serve to indicate relationships of nominals to each other (prepositions), to link parts of a discourse (connectives), to indicate whether entities and participants in a discourse are already identified or not (pronouns and articles), and to show whether they are close to the speaker or hearer (demonstratives). Frequently it can be shown that function words have their origins in content words. When a content word assumes the grammatical characteristics of a function word, the form is said to be "grammaticalized." Quite often what is grammaticalized is not a single content word but an entire construction that includes that word, as for example Old English *þa hwile þe* 'that time that' > *hwile* 'while' (a temporal connective).

1.2.1 A preliminary classification of grammatical forms

Not all grammatical forms are independent words. In most languages, at least some grammatical forms are bound as an affix or other category. Although there is no full agreement on definitions of grammatical forms, in general it is possible to speak of a continuum of *bonding* between forms that has a looser relationship between forms (i.e., independent words) at one end and a tighter relationship (i.e., grammatical affixes attached to stems) at the other. On this continuum there are various "cluster" or "focal areas" of the following nature (cf. Halliday 1961: 249; Bybee 1985; Hammond and Noonan 1988):

(a) Grammatical words with relative phonological and syntactic independence. For example, English prepositions can be found at the end of a clause without a noun phrase, as in *This is where we're at* and *This bed has been slept in*. In this position they have full segmental structure (unreduced vowels and consonants, e.g., [æt], not [ət]) and full prosodic structure (they can take stress).

(b) Derivational forms. Content words themselves often contain meaningful parts, known as derivational forms, that are neither inflections nor clitics

(see below). Many derivational forms add a meaning component without affecting the category in question. The *un-* of *unhappy* adds to the adjective *happy* the meaning 'not,' but does not change the adjectival status of the word. Similarly the *-ling* of *duckling* adds to the noun *duck* the new meaning 'young and small,' but does not change the nominal status of the word. Such derivational morphemes are part of the lexicon and can be called "lexical derivational morphemes." Other derivational forms do change the category of the word. For example, in the word *happily*, the suffix *-ly* derives an adverb from an adjective; in *swimmer*, the suffix *-er* derives a noun from the verb *swim*. Likewise, in the word *reclusive*, the suffix *-ive* derives an adjective from a noun. Because they not only add meaning but also serve to indicate grammatical categories, such "grammatical derivational morphemes" can be considered to serve a role between content and grammatical forms. Derivational morphemes are added to roots or stems, and the derived stems may be hosts for clitics and inflections.

(c) Clitics. These are forms that are not affixes, but are constrained to occurring next to an autonomous word, known as the host (for important treatments, see Klavans 1985; Zwicky 1985a; Halpern 1995). The diachronic process whereby a lexical form becomes a clitic is called "cliticization" (the corresponding verb is "cliticize"). The word clitic is a cover term for two varieties. A clitic that precedes the host is called a "proclitic," e.g., in colloquial English, *'s* in *'s* me 'it's me.' A clitic that follows its host is an "enclitic." Good examples of clitics in English are the *'m* in *I'm*, the *'re* in *you're*, the auxiliaries *'ll, 've* in *we'll, we've*, etc.; and discourse particles in many languages, e.g., in Latin, *-que* 'and':

(5) Conticuere omnes, intentique ora tenebant.
fell-silent all, intent-*que* gazes they-held
'All fell silent and intently held their gaze.' (*c.* 30–19 BC, Virgil, *Aeneid* II, 1)[2]

Clitics may be thought of as forms that are half-way between autonomous words and affixes (Jeffers and Zwicky 1980). They may share properties of both, although it is hard to make generalizations about which features will occur in a given instance. For example, clitics may resemble affixes in forming an accentual unit with the host. In Indonesian, where stress tends to occur on the next-to-last syllable of the word, the enclitic pronoun *nya* 'its' in *warná-nya* 'its colour' affects the stress in the host stem (contrast *wárna* 'colour'). On the other hand, clitics may behave more like independent words in having no effect on accent, as in Spanish *háblame* 'speak [sg.] to me!,' where the accent of the host *hábla* is unchanged by the extra syllable of the enclitic *me*.

(d) Inflections. These are always dependent and bound; that is to say, inflections by definition are always part of another word. Inflections reflect categories and properties of words such as gender, case, number, tense, aspect, and syntactic

relationships. In many languages, inflections are used to show agreement ("concord") in these properties or categories with some other word, e.g., English *this shoe* versus *these shoes*, where the forms of the demonstrative *this/these* reflect the singular/plural contrast in *shoe/shoes*.

1.2.2 Clines

Basic to work on grammaticalization is the concept of a "cline" (see Halliday 1961 for an early use of this term). From the point of view of change, forms do not shift abruptly from one category to another, but go through a series of small transitions, transitions that tend to be similar in type across languages. For example, a lexical noun like *back* that expresses a body part comes to stand for a spatial relationship in *in/at the back of*, and is susceptible to becoming an adverb, and perhaps eventually a preposition and even a case affix. Forms comparable to *back of* (*the house*) in English recur all over the world in different languages. The potential for change from lexical noun, to relational phrase, to adverb and preposition, and perhaps even to a case affix, is an example of what we mean by a cline.

The term "cline" is a metaphor for the empirical observation that cross-linguistically forms tend to undergo the same kinds of changes or have similar sets of relationships, in similar orders. "Cline" has both historical and synchronic implications. From a historical perspective, a cline is conceptualized as a natural "pathway" along which forms evolve, a schema which models the development of forms (see Andersen 2001). Synchronically a cline can be thought of as a "continuum": an arrangement of forms along an imaginary line at one end of which is a fuller form of some kind, perhaps "lexical," and at the opposite end a compacted and reduced form, perhaps "grammatical." Heine and his colleagues have suggested that the particular paths along which individual forms or groups of forms develop be called "grammaticalization channels" (see Lehman 1995[1982]) and the internal structure or relational patterns within these channels be called "grammaticalization chains" (Heine, Claudi, and Hünnemeyer 1991a: 222; Heine 1992). The metaphors "cline," "continuum," "pathway," "channel," and "chain" are to be understood as having certain focal points where phenomena may cluster. Most importantly, they are metaphors for labeling grammatical phenomena, not putative neurological or other elements of the language capacity.

The precise cluster points on the cline (i.e., the labels preposition, affix, etc.) are to a certain extent arbitrary. Linguists may not agree on what points to put on a cline, nor on how to define the cline in a given instance. They also may not agree on whether a particular form is to be placed in the lexical area or the grammatical area of the cline. But the relative positions on a cline are less subject to dispute.

For example, most linguists would agree that there is a "cline of grammaticality" of the following type:

content item > grammatical word > clitic > inflectional affix

Each item to the right is more clearly grammatical and less lexical than its partner to the left. Presented with such a cline, linguists would tend to agree that, in so far as they schematically reflect cross-linguistic generalizations, the points (labels) on the cline could not be arranged in a different order, although individual items may violate the order language-specifically (Andersen 2001). A number of such clines have been proposed, based on the many different dimensions of form and meaning that are found in language. Generally, they involve a unidirectional progression in bondedness, that is, in the degree of cohesion of adjacent forms that goes from loosest ("periphrasis") to tightest ("morphology").

It is often difficult to establish firm boundaries between the categories represented on clines, and indeed the study of grammaticalization has emerged in part out of a recognition of the general fluidity of so-called categories. It has also emerged out of recognition that a given form typically moves from a point on the left of the cline to a point further on the right, in other words, that there is a strong tendency toward *unidirectionality* in the history of individual forms. We will discuss unidirectionality and ways of conceptualizing the cline in some detail in Chapter 5.

1.2.3 Periphrasis versus affixation

Often the same categories can be expressed by forms at different places in the clines. Thus in English we have expressions that are "phrasal" or "periphrastic" (literally "occurring in a roundabout fashion") such as (6):

(6) a. have waited (perfect tense–aspect)
b. the household of the queen (possessive)
c. more interesting (comparative)

It is also possible to express tense–aspect, possession, and the comparative through affixes or changes internal to the stem word. In this case the categories are bound to a host and are said to be expressed "morphologically" or "affixally"as in (7):

(7) a. waited (past tense affixed *-ed*); sang (past tense signaled by internal change: contrast *sing*)
b. the receptionist's smile (possessive affix *-s*)
c. longer (comparative *-er*)

The distinction between the periphrastic and morphological expression of a category is important for the study of grammaticalization because of two diachronic

tendencies. One is for periphrastic constructions to coalesce over time and become morphological ones. While this and other tendencies are discussed in more detail later, especially in Chapter 6, a couple of examples follow:

(a) Definite nouns are marked in many European and other languages with an article that is separate from the noun, for example, English *the newspaper*, French *la rue* 'the street,' German *die Stadt* 'the city,' etc. In such languages definiteness is marked periphrastically (cf. English *the five yellow newspapers*, where the article is at some distance from the noun). But in some languages this sign of definiteness is an affix, which can usually be shown to derive from an earlier definite article or demonstrative. Thus in Istro-Romanian[3] the Latin demonstrative *ille* 'that' now appears as a suffix on nouns marking both definiteness and case, as in:

(8) gospodar-i-lor
boss-PL-DEF:GEN
'of the bosses'

Here *-i* marks plural and *-lor* is the definite genitive plural suffix deriving from Latin *illorum*, the masculine genitive plural of *ille*. Similarly in Danish, *-en* in *dreng-en* 'the boy' and *-et* in *hus-et* 'the house' are definite singular markers for common gender and neuter nouns respectively, and have their origin in earlier postposed demonstratives (cf. Old Norse *úlfr-inn* 'wolf-the' from **úlfr hinn* 'wolf-that'). In the modern languages they cannot be separated from the preceding stem.

(b) Various tenses and aspects of verbs are formed either with auxiliary verbs (i.e., periphrastic tense–aspect) or with verbal suffixes (i.e., morphological tense–aspect). Thus in Hindi the present tense is formed periphrastically by a verb stem plus the verb to *be:*

(9) mãĩ kursii par baiṭhaa hũũ.
I chair on sit: MASCSG be:1SG
'I sit on a chair.'

In Swahili, on the other hand, basic tenses such as the future are formed morphologically, with prefixes on the verb:

(10) Wa-ta-ni-uliza.
they-FUT-me-ask
'They will ask me.'

Morphological tense–aspect formations can often be shown to have developed out of earlier periphrastic ones. The Romance languages supply numerous examples of this, such as the Italian future *cantaremo* 'we will sing' or the French future *(nous) chanterons* from Latin *cantare habemus*, literally 'we have to sing.' We discuss this kind of development in the Romance languages in Section 3.3.1.