Hans Lindquist and
Christian Mair

Corpus Approaches to Grammaticalization in English

Studies in Corpus Linguistics

Corpus Approaches to Grammaticalization in English

Studies in Corpus Linguistics

Studies in Corpus Linguistics aims to provide insights into the way a corpus can be used, the type of findings that can be obtained, the possible applications of these findings as well as the theoretical changes that corpus work can bring into linguistics and language engineering. The main concern of SCL is to present findings based on, or related to, the cumulative effect of naturally occuring language and on the interpretation of frequency and distributional data.

Volume 13

Corpus Approaches to Grammaticalization in English
Edited by Hans Lindquist and Christian Mair

Corpus Approaches to Grammaticalization in English

Edited by

Hans Lindquist

Växjö University

Christian Mair

University of Freiburg

John Benjamins Publishing Company

Amsterdam/Philadelphia

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

Cover design: Françoise Berserik
Cover illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

Library of Congress Cataloging-in-Publication Data

Corpus approaches to grammaticalization in English / edited by Hans Lindquist, Christian Mair.
p. cm. (Studies in Corpus Linguistics, ISSN 1388–0373 ; v. 13)
"Contains revised versions of five out of six papers presented at the international symposium … organized by Hans Lindquist at Växjö University, Sweden, in 20–22 April 2001"--Pref.
Includes bibliographical references and indexes.
1. English language--Discourse analysis. 2. English language--Grammaticalization. I. Lindquist, Hans. II. Mair, Christian. III. Series.

PE1422.C67 2004
415-dc21 2004047741
ISBN 90 272 2284 3 (Eur.) / 1 58811 523 2 (US) (Hb; alk. paper)

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

Table of contents

Preface **vii**

Introduction **ix**
Hans Lindquist and Christian Mair

Three perspectives on grammaticalization: Lexico-grammar, corpora and historical sociolinguistics **1**
Terttu Nevalainen

Have to, gotta, must: Grammaticalisation, variation and specialization in English deontic modality **33**
Sali Tagliamonte

The semantic path from modality to aspect: *Be able to* in a cross-linguistic perspective **57**
Karin Aijmer

The passival and the progressive passive: A case study of layering in the English aspect and voice systems **79**
Marianne Hundt

Corpus linguistics and grammaticalisation theory: Statistics, frequencies, and beyond **121**
Christian Mair

Grammaticalisation from side to side: On the development of *beside(s)* **151**
Matti Rissanen

Are low-frequency complex prepositions grammaticalized? On the limits of corpus data — and the importance of intuition **171**
Sebastian Hoffmann

Life after degrammaticalisation: Plural *be* **211**
Laura Wright

Subject clitics in English: A case of degrammaticalization? 227
Laurel J. Brinton

Name index 257

Subject index 261

Preface

This volume contains revised versions of five out of six papers presented at the international symposium "Corpus research on grammaticalization in English" organized by Hans Lindquist at Växjö University, Sweden, 20–22 April, 2001. The original contributions have been supplemented by another four papers solicited from leading names in the field. In the invitation to the symposium, and later to the volume, the contributors were asked to present new empirical results on grammaticalization in English and to discuss methodological as well as theoretical questions in relation to these in the light of recent developments within the area. We believe that the result is an important collection of papers.

The original symposium was supported by The Bank of Sweden Tercentenary Foundation, for which we would like to express our gratitude. In the preparation of the present book we have benefited from valuable advice from Elizabeth Traugott and from an anonymous reviewer. Most of all, however, we would like to thank the participants in the symposium and the additional contributors for their efforts and their patience.

Växjö and Freiburg, December 2003

Hans Lindquist Christian Mair

Introduction

Hans Lindquist and Christian Mair
Växjö University / University of Freiburg

Research on grammaticalization has established itself as a major area in linguistic studies during the past two decades and grammaticalization is now both an important (but also controversial) concept in general and typological linguistics and a prominent type of explanation in historical linguistics. A noteworthy development towards the end of the nineties was a *rapprochement*, at least in English linguistics, between two scholarly communities which had previously tended to keep apart, namely corpus linguists and grammaticalization theorists. For historical corpus linguists, grammaticalization theory provided a welcome frame of orientation in their effort to analyse and systematize a fast-accumulating mass of data, and students of grammaticalization became increasingly aware of the potential of existing corpora and established corpus-linguistic methodology for their work. This was the dialogue which the 2001 symposium was intended to encourage, and which the present book is expected to continue.

There has been no shortage of publications on grammaticalization in recent years (cf., e.g., the collections of papers by Traugott and Heine (1991), Pagliuca (1994), Giacalone Ramat and Hopper (1998), Fischer, Rosenbach and Stein (2000), Bybee and Hopper (2001) and Wischer and Diewald (2002)). Apart from these collections of articles, a special issue of the journal *Language Sciences* (2001) was devoted to a critical review of the claims of grammaticalization studies. In their introduction, the editors, Campbell and Janda (Campbell and Janda 2001), raise a crucial question:

> Does the veritable flood of recent scholarship on grammaticalization correspond more closely to a "Great Leap Forward" or to "The Emperor's New Clothes"? That is, does it constitute a momentous advance in linguistic understanding or rest on an unfortunate misunderstanding?

In other words, is it the case that the semantic, syntactic and phonological processes that are subsumed under the term "grammaticalization" are already well-known and can be studied and explained without recourse to the concept (as Campbell and Janda seem to think), or is grammaticalization a substantive fact of linguistic evolution? Unlike Campbell and Janda, we would tend towards the latter view, pointing to the studies assembled in the present volume as corroborating evidence.

A strong trend in many studies of grammaticalization, including most of those mentioned above, is the focus on typological investigations and language universals and hence the use of data drawn from many languages. The present collection differs from such work in its concentration on one language, English. From the point of view of English linguistics this is natural, but we believe that in-depth studies of specific processes in the history of English may hold an important lesson for grammaticalization studies in general. There is now a corpus-linguistic working environment for English which comprises corpora which, taken together, are a body of evidence which is not only massive in size but also of extremely high quality because of its differentiated coverage of written and spoken material, of past and present stages of the language, and of different text types and genres. It is thus to be expected that grammaticalization processes can be studied in much more detail than would be possible for most other languages, and that the results thus obtained will lead to a refinement of the theoretical model. Specifically, the present volume pursues three aims. First, we make the general point that grammaticalization studies can gain from the systematic and principled use of large computerized corpora and the methods which have been developed within corpus linguistics. Second, and more specifically, we suggest necessary methodological refinements for grammaticalization theory, for example the need to incorporate the previously neglected factor of the differential speed of change in different textual genres. Third, we present a volume of in-depth studies of grammaticalization processes which are rigorously based on primary data, thus complementing work in general linguistics which is often concerned with languages for which only partial or sketchy documentation is available.

As the reader will see, all the contributions are therefore based on extensive use of various electronic corpora in the tradition of, for instance, the work on historical corpora carried out at Helsinki University (cf. Rissanen, Kytö, and Heikkonen 1997). At the same time the contributors aim to relate corpus practices to the recent theoretical concerns of grammaticalization studies. In the course of their papers, the authors deal with the following topics: grammaticalization and

historical sociolinguistics, lexicalization and grammaticalization, layering, frequency, grammaticalization and dialects, degrammaticalization and grammaticalization in a contrastive perspective.

In the first paper, *Terttu Nevalainen* relates grammaticalization research to the interface between lexis and grammar, lexico-grammar. Her paper also functions as a brief introduction to the field. Using historical data on adverbs, she claims that "[c]orpus linguistics is a means of observing grammaticalization in a way that provides an empirical methodology for investigating processes of language change in progress." The paper stresses the importance of diachronic data for distinguishing between register variation and ongoing change and also the important role of the extralinguistic embedding of the grammaticalization. Nevalainen concludes her paper with a detailed case study of the development of the lexical items *prett(il)y* and *fair(ly)*.

The English modal auxiliaries have often been adduced as typical examples of grammaticalization. Using the York English Corpus, *Sali Tagliamonte* studies variation in the expression of necessity and/or obligation. In contrast to suggestions made by Krug (Krug 1998; Krug 2000) that the *got*-construction is taking over as the marker of deontic modality in English, Tagliamonte shows that the change has not yet begun in the conservative York English. This dialect will therefore provide an interesting area of research in the future to see if, when and how *got to* and *gotta* take over.

Karin Aijmer's study differs from the others of this volume in that it is consistently contrastive. She looks at the expression of dynamic possibility and ability in English and Swedish, using the English-Swedish Parallel Corpus which contains Swedish originals and their English translations as well as English originals and their Swedish translations. She detects similar changes from ability to possibility, permission and actuality in the two languages, which indicates that there are rational, communicative causes for the changes.

In her paper, *Marianne Hundt* looks at the competition between the older passival form (*The house is building*) and the more recent grammaticalized progressive passive form (*The house is being built*) and the resulting situation of layering. Using corpus data she tests various claims and hypotheses from the literature about the two constructions. Her conclusion is that that there is presently stable rather than transitional layering of the two constructions, with the passival having become the marked patterning surviving especially with some verbs like *ship*, *show* and *play*.

In a paper discussing the relation between corpus linguistics and grammaticalization, *Christian Mair* asks himself the following questions: (1) What is the

role of discourse frequency?, (2) Can the use of corpora help us pinpoint incipient or ongoing grammaticalization? and (3) What are the insights into grammaticalization phenomena afforded by corpora that go beyond the (merely) statistical? Using the OED as a historical corpus (and arguing for the usefulness of this method) he is then able trace ongoing grammaticalization in a number of variable constructions by looking at changes in proportional frequencies, while at the same time emphasizing the importance of a qualitative analysis based on a theory of communication.

Matti Rissanen studies the historical background of the preposition, adverb and (now obsolete) subordinator *beside(s)*, illustrating its long-term semantic development and various stages of grammaticalization. His conclusion is that it patterns nicely with the development of a large number of adverbial connectives which have their roots in Old English and grammaticalized in Middle English.

The main concern in *Sebastian Hoffmann's* contribution is the role of frequency in grammaticalization and in particular whether low-frequency complex prepositions can be said to be grammaticalized. A complementary aim is to test the limits of corpus methodology. In order to do this he carries out a case study of the development of *in view of*, which he considers to have gone through generalization, subjectification, decategorialization and syntactic reanalysis. He stresses the importance of frequency in terms of relative proportion with respect to other realizations of the same context, which could be called "conceptual frequency" or salience. His claim is that "low-frequency combinations grammaticalize by analogy to their more frequent 'structural relatives'".

Laura Wright also underlines the importance of low-frequency items, which can transmit from generation to generation over long periods of time. Her paper deals with the fate of Middle English Southern third-person present tense plural *be*, which was replaced by Northern *are* during the 1500s. Using as material her own corpus from the Bridewell Court Minute books, she is able to show that *be* continued as a basilectal variant and was transported to the New World, where it became a factor in the development of African American Vernacular English. Wright argues that it was because *be* gained a new sociolinguistic property that it retained its grammatical function, "albeit in a regional dialect and at a minority ratio". The paper thus makes a contribution to the ongoing debate about the roots of AAVE.

In the final paper in the book *Laurel Brinton* counters a point of criticism that has been levelled against grammaticalization theory by Newmeyer (1998) and others, namely that the unidirectionality hypothesis does not hold and that various instances of decliticization can be brought forward as counter-examples

against it. Brinton's survey of the adduced instances of decliticization in various languages points to a vagueness in the definitions and a scarcity of unambiguous examples. Her own detailed empirical study deals mainly with the historical development of subject clitics in English but also with proclitics and object clitics. The diachronic evidence "suggests no universal loss or demise of cliticization of 2nd person subject forms, but rather a continued viability of subject enclisis and proclisis, as well as of object enclisis, in the history of English." The "return" of full forms is thus not decliticization or degrammaticalization, but rather a replacement by full forms which have existed in the language all along.

In their own ways, all nine papers show without a doubt that a synthesis of corpus methodology and grammaticalization studies leads to new and interesting insights about the mechanisms of language change and the communicative functions of language. One of the more specific gains is that we are now able to assess more precisely the role that frequency plays in grammaticalization. Frequency emerges as an interesting corollary of grammaticalization rather than as a primary cause, and some processes of grammaticalization do not seem to involve an increase in discourse frequency at all. Where it occurs, increasing discourse frequency is often a delayed sign of grammaticalization that has been going on for quite some time rather than a decisive factor in the process itself. A question worth asking is which kind of frequency one should look at: absolute discourse frequencies of grammaticalizing forms in corpora or proportional frequencies of lexical and grammatical uses? Diachronic corpora are indispensable in exploring these problems. Another thing which is also made very clear in this volume is that corpora can be (and should be) sources for qualitative as well as quantitative analyses.

We hope that the studies in this book will inspire further work on English and other languages for which large corpora are available. For languages without large corpora, the results might provide motivation for the creation of such corpora.

References

Bybee, J. and Hopper, P. J. (eds.) 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

Campbell, L. and Janda, R. 2001. "Introduction: Conceptions of grammaticalization and their problems". *Language Sciences* 23 (2–3): 93–112.

Fischer, O., Rosenbach, A. and Stein, D. 2000. *Pathways of Change: Grammaticalization in English.* (Studies in Language companion series 53.) Amsterdam and Philadelphia: John Benjamins.

Giacalone Ramat, A. and Hopper, P. J. 1998. *The Limits of Grammaticalization.* (Typological Studies in Language 37.) Philadelphia and Amsterdam: John Benjamins.

Krug, M. 1998. "*Gotta* — The tenth central modal in English? Social, stylistic and regional variation in the British National Corpus as evidence of ongoing grammaticalization". In *The Major Varieties of English: Papers From MAVEN 97, Växjö 20–22 November 1997.* H. Lindquist, S. Klintborg, M. Levin and M. Estling (eds.). Växjö: Acta Wexionensia.

Krug, M. 2000. *Emerging English Modals: A Corpus-based Study of Grammaticalization.* (Topics in English Linguistics 32). Berlin: Mouton de Gruyter.

Newmeyer, F. J. 1998. *Language Form and Language Function.* Cambridge, MA and London: MIT Press.

Pagliuca, W. 1994. *Perspectives on Grammaticalization.* (Current Issues in Linguistic Theory 109.) Philadelphia and Amsterdam: John Benjamins.

Rissanen, M., Kytö, M. and Heikkonen, K. (eds.). 1997. *Grammaticalization at Work: Studies of Long-term Developments in English.* (Topics in English Linguistics 24.) Berlin: Mouton de Gruyter.

Traugott, E. C. and Heine, B. (eds.). 1991. *Approaches to Grammaticalization.* 2 vols. (Typological Studies in Language 19.) Amsterdam and Philadelphia: John Benjamins.

Wischer, I. and Diewald, G. 2002. *New Reflections on Grammaticalization.* (Typological studies in language 49.) Amsterdam and Philadelphia: John Benjamins.

Three perspectives on grammaticalization

Lexico-grammar, corpora and historical sociolinguistics

Terttu Nevalainen
University of Helsinki

The intersection of lexis and grammar normally provides the focus for the study of grammaticalization. This paper discusses how the focus can be widened by corpus linguistic and sociolinguistic approaches, but it also illustrates the challenges that grammaticalization processes present to corpus-linguistic principles such as total accountability and normalcy.

The three perspectives on grammaticalization come together in a case study which raises the issue whether *-ly* adverbialization constitutes a process of lexicalization or grammaticalization. Historical corpus data on the intensifiers *fairly* and *pretty* drawn from Nevalainen and Rissanen (2002) provide evidence for two complementary scenarios. However, placed in a sociolinguistic context, processes of adverbialization prove no different from other forms of language change in that they may be highly susceptible to language-external conditioning.

1. Introduction

Grammaticalization can be approached from several vantage points. This paper begins at the macro level, and relates grammaticalization research to the interface between lexis and grammar, or *lexico-grammar*, for short. The second perspective that will be brought to bear on the study of grammaticalization is *corpus linguistics*, which in turn leads on to a third angle, that of *sociolinguistics*. While discussing some fundamental issues raised by these three perspectives, I will argue that valuable insights can be gained by considering all three of them in analysing, observing and contextualizing processes of grammaticalization. I will suggest that it is particularly the last two that have received too little

systematic attention in grammaticalization research. But since neither grammatical theory, corpus methodology nor sociolinguistics were primarily developed for the needs of grammaticalization research, processes of grammaticalization can also serve as testing ground for the degree to which they are capable of meeting the challenges presented by language change of this kind.

In *analysing* grammaticalization, the starting point is usually the cline or continuum between lexis and grammar, and the process is defined as one making lexical items grammatical, or existing grammatical items even more grammatical. Recent historical research has, however, focused on the problems inherent in basic linguistic categorization: problems may be encountered even with the distinction between adjectives and adverbs when they are examined in syntagmatic terms (Brinton 2002). I will enlarge on the problem with some adverb data that I suggest represent a weakly codified area in English grammar, hence permitting several alternative analyses.

Corpus linguistics is a means of *observing* grammaticalization in that it provides an empirical methodology for investigating processes of language change in progress. Problems are however created for corpus annotation by syntactic ambiguity, mergers of constructions, and the key issue in grammaticalization, gradience. The principles adopted to tackle these problems will decide the potential usefulness of tagged and parsed corpora for grammaticalization research. I will discuss the principles of total accountability and normalcy in more detail, and illustrate different solutions to the problems created by grammaticalization. My illustrative material comes from various historical and Present-day English corpora.

Sociolinguistic methods may be employed to *contextualize* processes of language change even further by focusing on their social embedding. This perspective is enabled by large, sociolinguistically varied corpora. Now the question is whether grammaticalization processes are different from other processes of language change in social terms. The research carried out within the framework of historical sociolinguistics suggests that the answer is in the negative (Nevalainen and Raumolin-Brunberg 2003). In other words, viewed from a sociolinguistic angle, the distinction between grammaticalization and other morphological and syntactic changes seems to disappear. If this is the case, then a more central position in grammaticalization research may have to be accorded to language external factors influencing the outcome of the process.

All these aspects of grammaticalization research will be brought together in my illustrative case study of adverbialization in the final sections of the paper. I will ask how processes of *-ly* adverbialization should be modeled, whether they

constitute processes of lexicalization or grammaticalization. I will test the hypotheses generated by both models against historical corpus data on the intensifiers *fairly* and *pretty* (Nevalainen and Rissanen 2002) and conclude by outlining two complementary scenarios to account for the data. The processes will also be placed in a wider sociolinguistic context which suggests that the outcome of processes of adverbialization may also be crucially susceptible to language-external conditioning.[1]

2. Approaching grammaticalization

My point of departure is the traditional contrast between grammar and lexis, which implies that linguistic items can be labelled as either grammatical or lexical. According to this view, grammatical items constitute closed classes whose membership is more or less fixed, while lexical items form open classes, which can be freely added to by means of word-formation processes. The various processes that make lexical items grammatical (or existing grammatical items even more grammatical) are therefore defined as *processes of grammaticalization*.[2]

An interesting consequence of this view of grammaticalization is that it necessarily breaks away from the rigid Saussurean dichotomy between *synchrony* and *diachrony*. Significantly, however, it is a historical linguist, Antoine Meillet, who is credited with having coined the term *grammaticalization* in the early 20th century, before the dichotomy itself came to dominate linguistic thinking. The fact that language change is always present in a living language may easily be forgotten or brushed aside by linguists who concentrate on the present-day "state of the language". This has been particularly true of most representatives of the 20th-century structuralist and mainstream generative traditions, but no less so of large corpus-based descriptive grammars of the latter half of the 20th century. Indeed, when synchronic variation is being described without access to diachronic data, it becomes increasingly difficult to distinguish between register variation and ongoing change. This is a point I shall return to in the empirical part of this paper.

By no means all synchronic linguists have ignored the grammatical prerequisites for change. Although few of them may address the issue of grammaticalization as such, there are those who have long argued that the grammar and lexis of a language like English do not constitute discrete categories. Edward Sapir (1921:38) put it simply by saying that "all grammars leak". M.A.K. Halliday refined the notion by suggesting that "the distinction between closed

system patterns and open set patterns in language is in fact a cline" (1961: 247), and that "formal patterns in all languages shade gradually from the grammatical into the lexical" (Halliday et al. 1964: 35).

Much to the same effect, Dwight Bolinger (1961) introduced the notion of *gradience*, Randolph Quirk (1965) the notion of *serial relationship*, and John Ross (1972) presented his concept of *squish*, such as the squish of nouniness. All these discussions are synchronic rather than diachronic in nature.

A great deal of work has also been done on the diachronic side, much of it inspired by the theoretical thinking of Elizabeth Traugott. Her 1982 model related grammaticalization directly to the ideational, interpersonal and textual components of grammar, the three building blocks of the systemic-functional grammar developed over the years by M.A.K. Halliday. Traugott argued that grammaticalization proceeds from ideational to interpersonal and textual functions and never reverses its course; hence the unidirectionality argument.

In the 1980s, synchronic grammaticalization research was boosted both by convergent interests of projects on typology and language universals, work by Christian Lehmann and others, and by individual 'non-mainstream' grammarians like Talmy Givón. Several volumes dedicated to the topic came out in the early part of the next decade combining synchronic and diachronic views. They include Paul Hopper and Elizabeth Traugott's *Grammaticalization* (1993) and the edited works by Heine, Claudi and Hünnemeyer (1991), and Traugott and Heine (1991). The methodology used in these studies was largely qualitative rather than primarily quantitative and corpus-based.

From the 1990s onwards, systematic corpus methods have steadily gained momentum particularly in diachronic research on grammaticalization. In 1997 the Helsinki team published a volume entitled *Grammaticalization at Work* (Rissanen et al., eds., 1997) which analysed such long-term processes of grammaticalization as *pronominalization* and *adverbialization* making quantitative generalizations on the basis of corpus evidence. Issues to do with adverbialization will reappear in the empirical sections of this paper as this is the subject of my case study, which presents both descriptive and theoretical challenges to grammaticalization research.

To round off this brief introductory survey, a few words can be said about sociolinguistics. Diachronic corpora such as the *Corpus of Early English Correspondence* have made it possible to approach the issue of language change, including processes of grammaticalization, from a sociolinguistic perspective. A sociolinguistic approach makes it possible to relate these processes to their social embedding across time. This real-time method was adopted by Minna

Palander-Collin (1999) in *Grammaticalization and Social Embedding: I Think and Methinks in Middle and Early Modern English*. The social profiles of processes of grammaticalization are also compared with other kinds of syntactic and morphological changes in Nevalainen and Raumolin-Brunberg (2003).

3. Three perspectives on grammaticalization

3.1 Grammar and grammaticalization

Today there appears to be a far-reaching consensus among linguists that grammatical and lexical categories constitute a cline. This continuum is regarded by Hopper and Traugott (1993: 7) as a *cline of grammaticality* that ranges from content items to inflexional affixes as follows:

> Cline of grammaticality:
> **content item > grammatical word > clitic > inflectional affix**

This view is only a rough and ready consensus model, and does not attach precise world-class labels to specific points on the cline. As such it does, however, suggest a basic divide between grammatical and content words. Halliday also recognized this when he wrote that "the distinction between closed system patterns and open set patterns in language is in fact a cline" but added that "the theory has to treat them as two distinct types of pattern requiring different categories" (Halliday 1961: 247).

For many linguists, the distinction is based on *paradigmatic* criteria such as whether the number of the terms in a category is finite or not, whether the terms are exclusive of all the others, and whether adding a new term to the system changes the meaning of all the others (Halliday 1961: 247). Criteria like these place pronouns at the more grammatical end of the cline than nouns, for instance, although the grammatical functions of the two categories remain much the same.

Things often become less clear-cut, however, if *syntagmatic* criteria are included in the discussion. Recent historical work on grammaticalization questions some fairly basic word-class distinctions such as that between adjectives and adverbs on syntagmatic grounds. Laurel Brinton (2002) discusses the grammatical status of temporal adverbs used adjectivally in Early and Late Modern English. They include adverbs with anterior, posterior and durative time reference, as in examples (1) to (3) *the now King* ('the present king'), *the then wife*, and *the since Earl* from (Brinton 2002: 80, 82).

(1) and the **now** King/The quondam Mounsieur shall not desire me this. (1653, Heminges, *The Fatal Contract*, v.ii.525–36)

(2) susanna martin the **then** wif of Georg martin being brought to Court for a wich (1692, *The Salem Witchcraft Papers*, Vol. 2, 572)

(3) eldest sonne of the **since** Earle of Norwiche (1620–1706, Evelyn, *The Diary of John Evelyn*, 28 July 1641)

Cases like these appear to be restricted to temporal and spatial adverbs. Manner adverbs, for instance, do not have the premodifier option. Brinton offers several possible interpretations for the process, which might be analysed as (a) word-formation (adverb > adjective); (b) lexicalization (grammatical > lexical); (c) grammaticalization (freely occurring adverb > restricted attributive modifier); or (d) degrammaticalization (more grammatical > less grammatical). On distributional grounds, the second alternative, lexicalization, looks like a fair candidate. If it is accepted, the first could also be argued for. Similarly, if (a) is the case, (d) follows. In this case, we might also consider (c), which attributes varying degrees of grammaticalization to functional and structural categories in the clause.

However, (c) could not be generalized across categories as it would entail, among other things, that all adjectives when used attributively are more grammatical than the same adjectives used in verbal complementation, and therefore more grammatical than adverbs, which by definition cannot function as attributive modifiers. This would clash with the more common notion that adverbs come closer to the grammatical end of the grammaticality cline than adjectives because, for instance, they are mostly derived from other word classes, and few of them can be inflected (see further Nevalainen 1997:180–182).

When the syntagmatic axis is considered, similar problems also appear with adverbs and adjectives other than temporal and spatial ones. Here the difficulty is similarly fundamental in that it hinges on basic word-class distinctions raising both the question whether some adverb-like words ought to be classified as adjectives in certain contexts on syntactic grounds and, conversely, whether words usually classified as adjectives should in some contexts also be labelled as adverbs. Let us begin by considering the two instances shown in examples (4) and (5).

(4) The flowers smell **sweet/sweetly**.

(5) He felt **bad/badly** about it.

These examples come from *The Comprehensive Grammar of the English Language* (Quirk et al. 1985:407–408). The authors analyse (4) and (5) as cases involving some uncertainty in the use of adverb and adjective forms with verbs

such as *smell*, *feel*, *look* and *sound* when they have the percept as subject. In other words, the case is viewed as one of indeterminacy in certain patterns of complementation, not in the word-class labels of the complements themselves (*sweet/sweetly*, *bad/badly*).

Historically, *-ly* forms with verbs of appearance are more widespread. Two examples from the *Helsinki Corpus* are reproduced in (6) and (7) (Nevalainen 1997: 150).

(6) As leene was his hors as is a rake, And he nas nat right fat, I undertake, But looked holwe, and therto **sobrely**.
(Chaucer: *The Canterbury Tales*, *The General Prologue* 289)

(7) it can receive no Light but at the Doors and Windows of the Porch, whereby it looks more **solemnly**; (Fryer: *A New Account of East India and Persia* I 186)

If we regard both *sobrely* and *solemnly* as adverbs, we have to explain the fact that an adverb can be conjoined with an adjective in (6) ('hollow and soberly'). Two alternative solutions to this problem have been suggested: Killie (2000a) opts for an adjective interpretation, suggesting that cases like this continue the by now recessive pattern of derivation of adjectives by means of *-ly*.

A competing analysis operating on the syntagmatic axis is offered by Gisborne (2000). He starts with the assumption that *sobrely* and *solemnly* in cases like (6) and (7) are indeed adverbs, but what has changed over time is their participation in constructions that occupy different positions on the continuum between complementation and adjunction. With verbs like *feel, look* and *smell* Gisborne identifies an intermediate attributary construction, which has the syntax of adjuncts but the semantics of predicative complements. It is this construction that relatively freely licenses adverbs as predicative complements in Middle and Early Modern English but only occasionally so in Present-day English. Gisborne (2000: 56) suggests that his approach is "consonant with most accounts of grammaticalization, where change is observed to be gradual and to exploit indeterminacy in syntax."[3]

A similar pattern of grammatical indeterminacy involving adjectives and adverbs, and complementation and adjunction, is found in (8) and (9), from Quirk et al. (1985: 1565).

(8) We stood **motionless**.

(9) He fell **flat**.

The analysis offered by the authors removes the adjunction alternative by viewing cases like (8) and (9) as instances where a secondary word-class change has taken place turning the intransitive verbs *stand* and *fall* into copulas, thus allowing *motionless* and *flat* to be treated as adjectives functioning as complements. This analysis solves the problem of having to decide on their word-class status, but it also builds an element of arbitrariness into the analysis. Regarded as a regular word-formation pattern, it stretches the copula category towards the lexical end of the grammaticality cline or, alternatively, shifts intransitive verbs towards the more grammatical end of it. Thinking of the basic distinction between grammatical and lexical elements, we might argue that this is not a typical case of word-formation — something the authors themselves also point out when they introduce the notion of secondary word-class changes (Quirk et al. 1985: 1564).

Comparing now the analyses proposed for examples (4)–(9) from the syntagmatic viewpoint, three different interpretations can be outlined, but only the second may be accounted for in terms of grammaticalization.

– no structural change: relic pattern (*look* + adjective in *-ly*)
– structural change: pattern loss (*look* + adverb in *-ly*)
– lexical change: conversion (intransitive verb → copula)

The first one, suggested by Killie, introduces an historical angle offering an etymological explanation of a pattern analysed nowadays as syntactic variation. In the second, Gisborne interprets the same data as a case where an attributary construction becomes streamlined according its semantic rather than syntactic characteristics, and in the course of time loses its ability to pattern with adverbs. Finally, the third case, taken from the Quirk grammar, suggests that an apparent distributional conflict between lexical forms and their functions (intransitive verb followed by an adjective) may in this case be solved by reinterpreting the lexical status of the verb.

In my earlier work, I viewed cases like these as belonging to a weakly codified area in the English grammar, something that has repercussions on the discussion of grammaticalization (Nevalainen 1997: 152). One of the problematic issues raised by cases like this is how to account for them in synchronic descriptions, assuming (a) that this kind of indeterminacy represents variation inherent in the grammars of living languages, but (b) that it may in the course of time result in more clear-cut grammatical, i.e. categorical, distinctions.

Much of the general discussion of the descriptive problem hinges on the scope of grammatical analysis. The distinction between grammatical and

content items suggested by the cline of grammaticality can obviously be refined, and has to do with the depth of descriptive detail, or, in Halliday's terms, *delicacy*. One of his definitions of delicacy runs as follows:

> A cline resembles a hierarchy in that it involves relation along a single dimension; but instead of being made up of a number of discrete terms a cline is a continuum carrying potentially infinite gradation. (Halliday 1961: 249)

This is usually understood in terms of gradience within the categories placed on the cline. Verbs, for instance, do not just divide into lexical or main verbs and grammatical or auxiliary verbs, but gradient distinctions can indeed be made between the two, including quasi-auxiliaries, copulas, other linking verbs, and so on. Following Bas Aarts (2004), we may refer to them as cases of *subsective gradience*, i.e., gradience within a grammatical category. By contrast, the gradience inherent in Gisborne's attributary constructions discussed above obtains between two categories, and would come under Aarts's *intersective* gradience. Both kinds are involved in the definition of grammaticalization processes cited in Section 2.

There are various ways of dealing with both subsective and intersective gradience, ranging from conflating grammatical categories to working with structural prototypes. In Hallidayan terms at the less delicate end, superordinate categories have a long history in English grammars. Otto Jespersen's traditional category of particles (1924: 87–90) covers adverbs, conjunctions, prepositions and interjections. More recently, Huddleston and Pullum (2002: 598–601, 612–617) fuse some homonymous adverbs, prepositions and conjunctions. Structural prototypes constitute the clear cases positively identified by the grammarian. A more delicate analysis is carried out when categories are described in terms of features matrixes. Features are employed by Quirk et al. (1985), for instance, to account for subsective gradience within word classes; Table 1 below illustrates this principle.

3.2 Corpus linguistics and grammaticalization

However gradience and processes of grammaticalization in progress are described in principle, they often pose empirical problems for corpus linguists who are tagging and parsing corpora for general use. Historical corpora are a case in point here. Two approaches to corpus annotation will be discussed below, those of *total accountability* and *normalcy*. But before going into them in more detail, it is worth noting that the systems of automatic part-of-speech

tagging developed for Modern English are usually quite successful and error rates can remain as low as four or five per cent (McEnery and Wilson 1996: 126). Still, as pointed out by Biber et al. (1998: 262), no automatic tagger is 100 per cent accurate.

Stig Johansson (1985: 208) introduces the goal of *total accountability* in corpus annotation in the context of the *Lancaster-Oslo/Bergen Corpus*. He specifies that total accountability requires that "a linguistic description should account for all the data in a body of texts selected in a systematic way". Three particular problem areas are found in post-editing word-class tagging: (1) ambiguity (adjective and adverb tags for items like *just*); (2) mergers of two structural interpretations (mono- and ditransitive readings for cases like *he bought her flowers*); and (3) gradience (prepositions v. prepositional adverbs: *from/down here*). Surprisingly many grammatical or closed-class items proved hard to classify in context (Johansson 1985: 214). *All*, for instance, can function as a pronoun, a determiner, and an adverb. It may form constructions such as (10) which can be analysed either as instances of a floating prequantifier ('all the spectators') or an intensifying adverb ('all in tears').

(10) The spectators were all in tears.

I will return to syntactic homonymy as a pathway that can be traced in grammaticalization in Section 4. It need not be a tagging problem if the functional categories in question can be sufficiently disambiguated and do not involve vagueness. In both cases, one solution would be to use *portmanteau tags*, assigning problem items multiple tags. The taggers of the *Lancaster-Oslo/Bergen Corpus* of written British English preferred disambiguation, however, and followed the *principle of normalcy*, according to which an ambiguous instance of a word was given the tag it had in the unambiguous cases.

Anchoring the analysis in prototypical instances is standard procedure in corpus annotation, but applying the principle across the board can hide variation and disappoint the corpus user who assumes a strong version of total accountability and expects annotation to show the 'correct' analysis. Grammaticalization presents particular annotation problems in corpora, not only historical but also modern, when it involves *univerbation*, i.e. the telescoping of syntactic phrases into one-word grammatical expressions. In Present-day English such instances include linguistic variables such as *gonna* v. *going to* and *innit* v. *isn't it* (Andersen 2001, Berglund 2000). A historical example is offered by the preposition and conjunction *because*, which goes back to the Middle English prepositional phrase *by cause of* and its variant forms. For the corpus

annotator the question is when the phrase ceases to be a phrase and ought to be tagged as a one-word unit, either as a preposition or conjunction.

The normalcy principle cannot be applied blindly when items like this are studied without previous baseline data, i.e. either case studies or general descriptions of the period in question. Helena Raumolin-Brunberg (1991:98) illustrates this issue with a feature matrix of complex prepositions in her study of the NP structures of Sir Thomas More, an early 16th-century statesman and humanist. What we find is gradience along a cline ranging from prepositional phrases to more fixed expressions. Table 1 reproduces her data for *because of* and *on the behalfe of* showing that *because of* had five of the nine invariant features posited for a complex preposition, but *on the behalfe of* only three. *Because of* had advanced further towards the fixed end of the cline, and the author therefore decided to treat it as a complex preposition but continued to analyse *on the behalfe of* as a prepositional phrase.

Table 1. Complex prepositions in Sir Thomas More's language (based on Raumolin-Brunberg 1991:98)

Criterion/Item (Prep1+N+Prep2)	***by cause of/because of***	***on the behalfe of***
1. Prep2 may vary	−	−
2. Number of N may vary (sg./pl.)	−	−
3. Determiner of N may vary	+	+
4. Prep1 may vary	+	+
5. Prep2+complement may be replaced by a possessive pronoun	+	+
6. Prep2+complement may be omitted	−	−
7. Prep2+complement may be replaced by a demonstrative	−	+
8. N may be replaced by a related N	+	+
9. Words always spelled separately	−	+

When tagging a corpus, a principled solution of this kind is not available without prior research into the material. A standard solution is to provide the corpus user with a comprehensive lexicon which specifies the part-of-speech coding applied to all grammatical and lexical items or strings of items. With some earlier periods of the language, Middle English in particular, full part-of-speech annotation and disambiguation has not, however, always been feasible because of the wide-ranging spelling variation in the language at the time. This obstacle has been felt particularly acutely by the pioneers who have produced

not only tagged but also parsed corpora of the period such as the *Parsed Penn-Helsinki Corpus of Middle English* (PPCME), covering the period from 1150 to 1500.[4] The manual to the first edition of the corpus (1994) states:

> The nature and extent of spelling variation in Middle English has imposed some limitations on the corpus. Because of this variation, for instance, we have not tagged each word for part of speech. This decision was forced on us by the tremendous overlap in spelling among words (the spelling 'here', for instance, is used to represent four different parts of speech: noun, verb, adverb and personal pronoun). Since function words are especially prone to overlap and bear much of the load in tagging, automatic tagging, which can be done with high accuracy for Modern English, is not currently feasible in our corpus, and hand tagging would have been too time consuming.

These additional problems for corpus annotators posed by historical data mean another step away from total accountability in the sense outlined by Johansson (1985). The partial-annotation solution adopted in the first edition of the PPCME was justified by the time and funding available for annotating a large number of texts in their original spelling format, and served as an intermediate stage in the annotation history of the corpus. Multiple processing with total accountability as the ultimate aim may be one way to proceed with general-purpose corpora of earlier historical periods such as Middle English.

Besides a staggering amount of spelling variation, homonymy in Middle English is created by *morphological levelling*. This process presents an interesting problem area for historical corpora. Here the PPCME applies the principle of normalcy, no case marking at all being encoded for Middle English. In many instances, indicating levelled case endings would have involved the use of portmanteau tags or called for difficult and arbitrary judgements concerning the status of certain spelling conventions such as the final ⟨e⟩. Nevertheless, from the point of view of grammaticalization research, this means that we do not have automatic access to data reflecting the loss of inflections and grammatical gender in early Middle English, which constitute a massive process of grammatical levelling or, as one might call it, *degrammaticalization* (see Jones 1988).[5]

A very different approach to these problems has been adopted by those historical corpus linguists who have decided to opt for *normalizing* their texts. Normalization can range from disambiguating homonyms down to modernization of the morphology and syntax of the text. An example of supplying homonyms with portmanteau tags is given by Manfred Markus (2000: 192), and shown in part in (11). The passage from the *Ayenbite of Inwyt* amply illustrates the Middle English range of homonymy, including forms of the definite article

and the demonstrative pronoun *that*, and the singular and plural of the demonstrative *this*. It is no coincidence that the demonstrative and the definite article should have the same spelling — the latter is the result of a process of grammaticalization based on the former. But to make the normalized glosses (indicated by $N and italicized) user-friendly, contextual disambiguation would be called for in the subsequent versions of the normalized text.

(11) $I þis boc is ywrite /
$N this/these book is ywrite /
$I uor englisse men, þet hi wyte /
$N for englisse men, the/that she/they/her wyte /
$I hou hi ssolle ham-zelue ssriue,
$N how she/they/her shall themselves shrive(n),
$I and maki ham klene / ine þise liue.
$N and make them clean / in this/these live/life.
$I þis boc hatte huo þet writ /
$N this/these book hatte who the/that writ /
$I AYENBITE OF INWYT.
$N *AYENBITE OF CONSCIENCE.*

3.3 Sociolinguistics and grammaticalization

Corpora not only provide the researcher with quantitative linguistic data on processes of grammaticalization but, increasingly, with data on their extra-linguistic embedding. Needless to say, it is stylistically and socially stratified corpora that will reveal the external complexity of on-going processes in the varieties and genres of Present-day English.

The same goes for diachronic processes of grammaticalization and their reconstruction. Some historical linguists would argue that, once under way, a linguistic change will follow a constant rate in all relevant linguistic environments. But even these proponents of the *constant-rate-hypothesis* do not suggest that there is something like an internal clock followed by linguistic changes. In Anthony Kroch's words (2001: 721):

> Nothing in the grammatical system undergoing change accounts for the rate of the change or for the fact that the change actually goes to completion rather than stalling or even reversing.

In order to be able to account for the time-courses of linguistic changes, we need real-time corpora. Multigenre diachronic corpora will enable the researcher to

tell whether, for instance, a given process is more likely to have had its origins in speech than in writing. Studies of longitudinal multigenre corpora such as the *Helsinki Corpus of English Texts* and the *ARCHER Corpus* of historical English registers systematically indicate that many complex prepositions and conjunctions such as the conditional subordinator *provided that* typically go back to the written mode of the language — their favoured register-orientation even today (Rissanen 2000: 126–127).

Most linguists would agree that the linguistic and external courses of language changes are not causally connected. We may therefore ask whether, in grammaticalization processes, the *social embedding* of the change can be shown to be in any way different from other processes of linguistic change. Jack Chambers (1995: 51) tackles the issue when he suggests that grammatical variables mark social stratification even more strongly than phonological variables. There is evidence showing, for instance, that some stable sociolinguistic variables such as negative concord stratify women more sharply than men.

These observations apply to grammatical variation in general, however, and have nothing to say about processes of grammaticalization as such. Within the framework of our *Sociolinguistics and Language History Project*, Helena Raumolin-Brunberg and I have studied a number of grammatical changes in Late Middle and Early Modern English using the *Corpus of Early English Correspondence* (CEEC; Nevalainen and Raumolin-Brunberg 2003). Some of them consist of 'mere' morpheme replacements in that a form originally used in a given regional or social dialect is generalized in the literate social ranks throughout the country. Others constitute processes of grammaticalization, especially those in which grammatical morphemes become more grammatical by acquiring new grammatical functions. One case is the transfer into the subject function of the object pronoun *you*, which ousted the historical form *ye* (example (12)). Another is the extension of the numeral *one* into pronominal and prop-word functions (example (13)), while a third is the replacement of *of*-phrases (example (14a)) by zero forms (14b) in the object of the gerund.

(12) a. … glad of Ratchelle's amendement. I praie you comme not moche at her yourself, leest **ye** get the dysease yourself. (John Johnson, 1545; Johnson, 395)

b. Mr. Douse is nowe at London for the same mater: if **you** spake with hym, **you** shall knowe all. (Sabine Johnson, 1545; Johnson, 245)

(13) I with all your lytell **ons** be in helthe, the Lord be prasid, (Sabine Johnson, 1546; Johnson, 668)

(14) a. as you have done by contynuall **charging of monney** (Ambrose Saunders, 1552; Johnson, 1610)
b. I promis myselfe the contentment of **meeting you**; (Lucy Russell, 1614; Cornwallis, 23)

Our conclusion so far is that, however defined, grammaticalization does not follow any one particular pattern of social embedding compared with other processes of supralocalization that we have investigated. The latter include the replacement of the southern third-person indicative present-tense suffix *-th* by the originally northern *-s*, and the generalization of the relative pronoun *which* and the disappearance of *the which*. See examples (15) and (16).

(15) a. This **hath** changed my resolucon, for I intend now to staie here a fortnight longer … (Thomas Wentworth, 1624; Wentworth, 215)
b. I haue sent you a partriche pye, which **has** the two pea chikeins in it … (Brilliana Harley, 1627; Harley, 3)

(16) a. Your letter by Mr. Brudenell I have receyved, for **the which** I thancke you. (Sabine Johnson, 1545; Johnson, 266)
b. Victor hath gotten a byll of him **which** is signed by hys aunt, Joes Diricken's wyfe, … (John Johnson, 1545; Johnson, 323)

Our answer to the question of the social *eigenart* of grammaticalization processes would thus provisionally appear to be in the negative. However, I will return to this issue in the case study in Section 4 with some words of caution. At this point I would simply like to suggest that so little is known about the social side of real-time processes of language change in general and grammaticalization processes in particular that only some very broad empirical generalizations may be advanced, such as the social and regional origins and gender affiliation of changes that spread throughout the country in the periods studied (Nevalainen 2000, Nevalainen and Raumolin-Brunberg 2003).

Although historical sociolinguistic work on grammaticalization and other linguistic changes in general is still in its fact-finding stage and has a long way to go to meet the requirements of *total accountability*, some points may be made concerning a theoretical issue that has been raised in recent grammaticalization literature. Janda (2001:304–315) notes that historical sociolinguistic research will need to throw more light on the basic tenet that language transmission, and hence language change, is discontinuous, and has to be acquired anew by each successive generation. What he has in mind is the basic sociolinguistic distinction between generational and communal changes as applied to grammaticalization. In *generational change*, grammaticalization proceeds by generational

steps, and people do not change their language after adolescence; in *communal change*, the reverse is the case, people change their language even in adulthood, and grammaticalization can proceed simultaneously across generations in the speech community (for the terms, see also Labov 1994: 83–84).

A compromise is suggested by our data in most cases, both for simple changes such as the replacement of *-th* by *-s* and changes more directly involving grammaticalization, such as the generalization of the object form *you* as a subject pronoun. When cross-tabulated generationally by apparent time and communally by real time, almost all the changes considered, regardless of their grammaticalization status, progressed on both axes (Raumolin-Brunberg 1996, Nevalainen and Raumolin-Brunberg 2003, ch. 5). This real-time historical sociolinguistic evidence could be taken to strengthen the point made above that the distinction between processes of grammaticalization and other changes taking place in grammar may be neutralized when they are diffused throughout the language community.

4. A case study: Adverbialization

4.1 Lexicalization or grammaticalization?

The case study I have selected for illustrating the points raised in Section 3 deals with *adverbialization*. At first sight, adverbialization might not look like a prime candidate for the purpose: as adverbs constitute an open class that can be freely expanded, they may be considered content words. Moreover, as shown in 3.1, above, it may not always be easy to distinguish between adjectival and adverbial functions on the basis of morphological form.

We can however position adjectives and adverbs relative to each other on the cline of grammaticality. A comparison like this is made by Elizabeth Traugott (1988: 134–136), who argues that, because adverbs are derived from other word classes, they come closer to the grammatical end of the cline than adjectives. The suffix *-ly* is typically used to derive adverbs from adjectives in Modern English. Another category feature that makes adverbs atypical content words is that few of them can be inflected for comparison.

However, since *-ly* suffixation is so common we could make the issue morphological and ask whether adverbialization really forms part of the *derivational* system of English at all. Perhaps it ought to be included in the *inflectional* system of the language instead. It fulfils the two major criteria set for inflectional processes.

The *-ly* suffix itself has minimal semantic content, and so meets the requirement of generality associated with inflectional elements. Furthermore, its productivity is almost unrestricted: it is possible to form adverbs from practically all adjectives by means of the *-ly* suffix in Present-day English.

As shown by Kristin Killie (2000b), it has become increasingly common to form *-ly* adverbs even from stative adjectives such as colour terms. Examples such as (17) from the Chadwyck-Healey LION database begin to appear from the 19th century onwards (there are 15 instances of the adverb *redly* in the LION nineteenth-century English fiction database as opposed to none in the previous centuries).

(17) …the westering sun was gleaming **redly** on the old hall, (Anne Brontë, *The Tenant of Wildfell Hall* (1848), Vol. 1, Ch. XV)

Another source of evidence to suggest the increasing productivity of the suffix in Modern English is sentence adverbs formed with *-ly* to express the speaker's point of view. This category of adverbs has become greatly diversified since Old English. As shown by Swan (1997), certain classes such as *evaluative* adverbs (called *style disjuncts* in Quirk et al. 1985) have only gained momentum in the last hundred years or so. They are illustrated by *naturally* in (18) taken from the Chadwyck-Healey 19th-century fiction database.

(18) So there was no avoidance; we had to tap it and taste it, and give a sup to the fellow who brought us the keg, and drink the health of the Captain. One could not be churlish; and, **naturally**, I could not abstain from letting Jamie try the spirit. (Sabine Baring-Gould, *The Roar of the Sea* (1892), Vol. 1, Ch. IX)

4.2 Modelling *-ly* adverbialization

As the high degree of morphological productivity of *-ly* is coupled with a specific grammatical role, some linguists indeed regard *-ly* suffixation as an inflectional rather than a derivational process (Marchand 1969, Baayen and Renouf 1996; also Quirk et al. 1985: 1556). Accordingly, *-ly* adverbs are taken to constitute a grammatical category, not a lexical one.

But as noted above, this high degree of productivity has not always been characteristic of *-ly* adverbialization. English still has suffixless adverbs, and stative and evaluative adverbs in *-ly* are a relatively recent phenomenon. To appreciate the modelling problem more fully in its historical context, let us return to clines. Besides the cline of grammaticality, where lexical words are

contained as a single unanalysed category (see 3.1), another cline may be posited for lexicality. In Hopper and Traugott (1993: 7) it ranges from a syntatic phrase to a derivational affix:

> Cline of lexicality:
> **syntactic phrase > compound > derivational affix**
> a basket *full* (of eggs) > a cup*ful* (of water) > hope*ful*

This development from a compounding element to a derivational suffix was also undergone by the adverbial suffix *-ly*. Etymologically it goes back to a noun denoting 'body', 'likeness', cognate with Scandinavian *lik*, and Present-day English *like*, for instance. The lexeme first formed *compounds* with nouns, and was later reduced to a derivational suffix to form adjectives. This was the case in Old English, in which *-lych* was used to derive adjectives from both nouns and other adjectives; we still have *-ly* derived adjectives such as *godly* and *kindly*. The Old English adverbial suffix was *-e*, which could be attached to all adjectives, both in *-lych* and to base forms, to form adverbs. In Middle English, the final *-e* was eroded like so many other weak suffixes, and the adverbial suffix was renalysed as *-lyche*, i.e. containing the adjectival suffix *-lych* and the earlier adverb suffix *-e*. This combination yielded the shorter *-ly* in Late Middle English (see Nevalainen 1997).

The adjective suffix *-ly* clearly has lexical origins, but in order to claim that the adverbial suffix *-ly* is inflectional and not derivational, we would need to posit a mechanism that would make the connection. Separating lexical and grammatical clines the way we have done above does not help one see how the transition from one to the other could take place on the paradigmatic plane. Combining the clines of grammaticality and lexicality into a sequence as shown below in Model 1 makes this clear. The derivational and inflectional suffixes do not meet.

Model 1
Processes on the lexical plane [**syntactic phrase > compound > derivational affix**] >||> processes on the grammatical plane [**content item > grammatical word > clitic > inflectional affix**]

If we wish to bridge the distance between the content and grammatical planes in morphological terms, we need a model showing a continuum between derivation and inflection. One such model has been suggested by Joan Bybee (1985) on the basis of a cross-linguistic survey of fifty languages. She bases her continuum on Sapir's distinction (1921) between material and relational concepts, using these semantic terms to construct her scale ranging from lexical to

syntactic elements. In her model, shown below as Model 2, semantic elements can be combined in different ways and these combinations placed on a continuum from the most highly fused means of expression, i.e. lexical expression, to the most loosely combined means, i.e. syntactic expression (Bybee 1985: 12). In Model 2, the degree of fusion decreases the closer to the syntactic end we get.

Model 2
lexical – derivational – inflectional – free grammatical – syntactic

This model would have no difficulty accounting for *-ly* adverbialization as a more syntactic process than, say, adjective formation by any derivational means. The adverb suffix regularly follows the adjective form, both base and derived, without changing its lexical meaning. Model 2 also captures the fact that in order for lexical material to grammaticalize it must be part of a syntactic expression. *Because* does go back to the noun *cause* but only when it appears in the prepositional phrase *by cause of*. The model hence supports Hopper's view (1991) of 'layering' in grammaticalization: the most recent layer in grammar consists of syntactic expressions, as is the case with periphrastic expressions of tense relations in English, and the oldest layer contains fused expressions such as the ablaut tense forms of strong verbs. If the model is applied to adjective and adverb formation by means of *-ly(che)*, a similar order is attested, the adverb suffix being built on the adjective suffix.

4.3 A case study of intensifiers

We have already seen how *-ly* adverbialization has gained ground in the history of English. However, before we can argue for the grammatical status of English adverb formation we shall have to consider *-ly* adverbs in Bybee's terms both as relatively loose expressions and as more fused grammaticalized or lexicalized expressions which need to be recorded in dictionaries. There is another common means of adverb formation that ought to be considered in relation to *-ly*, i.e. zero-formation. Model 2 predicts that this synthetic means represents an older layer in the English adverb system than *-ly* formation and is less productive than the more analytic pattern.[6]

The status of English 'zero-derived' adverbs has changed over the centuries. Some of them, such as *hard* and *soft*, go back to Old English, where they were formed by means of the suffix *-e*. A number of new adverbs were formed in Middle English by means of zero-derivation, as Donner's study of the *Middle English Dictionary* (1991) shows. My own investigation of the *Helsinki Corpus*

indicates, however, that the number of dual adverbs with both suffixed and suffixless forms was reduced in the course of the Early Modern English period (Nevalainen 1997). In Late Modern English, the use of zero-forms was typically condemned by prescriptive grammarians, particularly when it occurred in intensifiers. Robert Lowth (1762/1775: 93), for instance, proscribes collocations such *excellent well*, *extreme elaborate*, and *marvellous graceful* (for a listing of prescriptive comments, see Sundby et al. 1991: 200–203).

In order to be able to test semantically like with like, I will compare the adverbialization histories of two near-synonyms in Present-day English, the downtoners, or *moderators*, *pretty* and *fairly*, which both function as word-modifiers (for the term, see Paradis 1997: 27). They can both be used to modify adjectives in the sense 'moderately', but morphologically *pretty* as a zero-derived adverb is more fused than *fairly*.

4.3.1 *Is* ***pretty*** *historically earlier than* ***fairly?***

Model 2 suggests several hypotheses which might predict the relative time courses of the two word-modifiers. We could assume that (1) the modifier *pretty* is the earlier of the two. More generally, it could be assumed that (2) its over-all distribution as an adverb is now more restricted than that of *fairly*. It could also be conjectured that (3) both *fair* and *prettily* must exist as adverbs.[7]

Tables 2 to 5 (from Nevalainen and Rissanen 2002: 369, 370, 375, 377) confirm these hypotheses. The tables trace the history of *prett(il)y* and *fair(ly)* from the 16th century up to the present-day on the basis of diachronic corpora making a distinction between their adjective and adverb functions. In functional terms the adverbs are divided into modifiers and adverbials. Adverbials are clause-level constituents, while modifiers operate within the clause-constituent level. Intensifiers typically modify adjectives and other adverbs.

Table 2 indicates that the *-ly* adverb *prettily* did exist but was rare in Early Modern English. It was used as an adverbial, but one instance of a modifier function was also found, *pretily well*, shown in example (19). The adverb *pretty*

Table 2. The adjectival and adverbial occurrences of **prett(il)y** in the Early Modern English part of the Helsinki Corpus (HC; absolute figures)

	Adj.	**Adv. *pretty***	**Adv. *prettily***
E1 (1500–1570)	15	–	1
E2 (1570–1640)	5	5	1
E3 (1640–1710)	23	31	1

Table 3. The adjectival and adverbial occurrences of **fair**(**ly**) in the Early Modern English part of the Helsinki Corpus (absolute figures)

	Adj.	**Adv. (*-e*)**	**Adv. *-ly***
E1 (1500–1570)	66	7	–
E2 (1570–1640)	67	1	1
E3 (1640–1710)	44	–	3

is first attested as a modifier in the 16th century and gains ground in the course of the Early Modern English period; see examples (20) and (21).[8]

(19) *Sam.* Truely no, I thanke God I haue had my health **pretily** well, but yet me thinke my meate doth me no good of late. (HC; 1593 Gifford, *Handbook of Witches* A4R, GIFFORD)

(20) Boccace is **prettie** hard, yet understood: Petrarche harder but explaned. (1598 Florio *Dict.*, OED, s.v. *pretty* adv. 1a)

(21) I found certain very small, black, but glistering Spots of a movable Substance, each of which examining with my *Miscrocope*, I found to be a small round *Globule*; some of which, as they looked **prety** small, so did they from their Surface yield a very bright and strong reflection on that side which was next the Light; and each look'd almost like a **prety** bright Iron-Ball, whose Surface was **prety** regular, such as is represented by the Figure A. (HC; 1665 Hooke, *Micrographia* 13.5,44, HOOKE)

As shown by Table 3, the story of *fair* and *fairly* presents almost a mirror image to that of *pretty*. As predicted, the zero-form *fair* does exist but becomes rare in the course of the Early Modern period, although it does not disappear completely. It was used as an adverbial, as in (22), and very rarely as a modifier (once in the *Lampeter Corpus of Early Modern English Tracts*, 1640–1740).

(22) Hodge when I speake so **faire**: wilt stil say me nay: (HC; 1575 Stevenson, *Gammer Gurton's Needle* 69 Stevenso)

The *-ly* form *fairly*, which goes back to Old English, functions as an adverbial, as in (23). It is also occasionally used as a modifier, as in (24), but the modifier function does not become frequent until the late nineteenth century, as the *ARCHER* data in Table 4 shows.

(23) you are more dishonest then the Common women, for they **fairly** tell us what they are. (1690, Crowne, John, *The English Frier*, LION, Drama)

Table 4. *Fairly* in the ARCHER Corpus (1650–1990)

	Modifier	Adverbial
1650–1800	–	15
1801–1900	3	18
1901–1990	22	6

Table 5. *Fairly* and *pretty* as downtoning modifiers (moderators) and in other uses in the Lancaster-Oslo/Bergen, Freiburg-Lancaster-Oslo/Bergen, Brown and Freiburg-Brown corpora

	fairly		*pretty*	
	Modifier	Adverbial	Modifier	Adjective
LOB	75	8	57	47
FLOB	53	7	50	41
Brown	58	8	63	41
Frown	42	5	63	39

(24) He […] swallowed a whole dispensary of bolusses, draughts and apozems, by which means he became **fairly** delirious in three days, and so untractable, that he could be no longer managed according to rule; (1753, Smollett, Tobias, *Ferdinand Count Fathom*, LION, Fiction)

In the 20th century, the modifier *fairly* by far outnumbers the adverbial use of *fairly*. The late 20th century situation is shown in Table 5. Both *pretty* and *fairly* can be used side by side as downtoning modifiers in the sense 'moderately', but *fair* and *prettily* no longer have the modifier function.

4.3.2 *Adverb–adjective correspondences*

On the basis of constituency, the grammatical role of an adverb may be simply described at the phrase and clause levels corresponding to the modifier and adverbial functions, respectively. But as suggested above, more complex syntactic–semantic accounts can be combined with these basic syntactic notions. In the case of adverbs this has however proved less straightforward than, say, with pronouns because adverb meanings cannot always be deduced from the meanings of the corresponding adjectives. This complication suggests one of two things:

1. not all adverbs are directly adverbialized from the corresponding adjectives, or
2. adjectival senses they correspond to have once been attested but no longer exist.

If the first alternative is true, we have compelling evidence to suggest that adverbialization is not simply an inflectional process. In the second case, it would be easier to argue that adverbialization is an inflectional process, akin to such fully productive processes as tense-marking by means of the *-ed* suffix.

Pretty adverbialized as an intensifying modifier following the Middle and Early Modern English pattern of zero-formation. As an intensifier it is associated with the adjectival senses of *pretty*, 'fair-sized', 'considerable', and 'great', which introduced an element of quantification to the nouns they modified. The semantic change required of the scalar adverb *pretty* had already taken place in the adjective. The intensifier *pretty* therefore looks like a case of adverbialization from the adjective *pretty*. The fuzzy meaning of the adjective base correlates with the booster and moderator readings of the intensifier that appear in Early Modern English (ex. (21)).

The adverb *fairly* is rather more complicated in that the corresponding Old English adverb *fægere* already had a number of senses shared by the longer adverb. The *Oxford English Dictionary* suggests that the adverbs *fair* and *fairly* shared meanings such as 'equitably', 'honestly', 'impartially' and 'justly'; 'according to rule', 'in a proper manner' and 'becomingly'; 'moderately', 'not excessively', 'completely', 'fully', 'quite', 'clearly', 'distinctly' and 'plainly'. Most of the more abstract senses were first attested with *fair*, which was the more common of the two adverbs in Old and Middle English. The adjective *fair* does not seem to have displayed all these more abstract senses. As both adverbs coexisted for centuries, it is perhaps legitimate to find semantic links between them. If this interpretation is correct, we have evidence both for adverbialization from adjectives and for more adverb-related semantic developments.

Fairly continues to display considerable polysemy even today. It assumes a range of adverbial functions other than that of the moderator shown in example (i). Some of them are illustrated by examples (ii) to (iv) (from Nevalainen and Rissanen 2002: 361–362).

i. **intensifier: moderator**, 'pretty', 'quite':
   It would *fairly* soon be light and he would get a sense of direction again. (Bäcklund 1973: 147)
ii. **intensifier: maximizer**, 'completely':
iia. The general was *fairly* vanquished. (Bäcklund 1973: 144)
iib. He *fairly* rocketed past us on his motorbike. (*Longman Dictionary of Contemporary English*; s.v. *fairly* 3 infml. 'completely')
iii. **emphasizer** of truth value, esp. BrE, 'absolutely':
   In her anger, she *fairly* screamed at him. (Quirk et al. 1985: 585)

Even as an intensifier, *fairly* may not only serve as a downtoner but also amplify the meaning of the element it modifies ('completely', 'fully'). Although this maximizer use is recorded in dictionaries, it is archaic, as in (iia), or hard to distinguish from those associated with emphasis, as in (iib). An emphasizer proper is illustrated in (iii). Emphasizers reinforce the truth value of the clause or part of the clause they apply to, and can be related to modality distinctions expressed at the clause level. They are therefore classified as adverbial *subjuncts* rather than modifiers by Quirk et al. (1985: 583). In their grammar, subjuncts are allotted a syntactically subordinate role in comparison with other sentence elements, and fall in between modifiers and adverbial adjuncts.

To add to their syntactic homonymy, polysemous *-ly* items such as *fairly* typically also assume clause-level functions as adverbials and serve as manner *adjuncts*, as in (iva). Since *fairly* is also capable of assuming subject-oriented readings, (ivb) may be paraphrased as 'I'm fair/honest in citing it'. Subject-oriented adverbials come under item subjuncts in Quirk et al. (1985: 572). However, example (ivb) might equally be analysed as containing a sentence adverbial that conveys the speaker's comment on the message content, which could then be paraphrased as 'my action of citing it is fair/justified'. Adverbials like this fall outside the sentence structure proper and are classified as *disjuncts* by Quirk et al. (1985: 621), corresponding to Swan's evaluative adverbs.

iv. **manner adverbial**, 'beautifully'; 'neatly'; 'rightly'

iva. You excelled in writing *fairly.* (*Webster's Third New International Dictionary;* s.v. *fairly* 1a(2))

ivb. But I can *fairly* cite it as a case in point of what I am saying. (Bäcklund 1973: 144)

But rather than copy all these syntactic-semantic features from the shorter adverb, *fairly* appears to develop them in a given order. As suggested by the historical corpus data discussed in greater detail in Nevalainen and Rissanen (2002), its adverbial functions appear before its modifier functions. The adverbial functions are followed by the truth emphasizer and amplifier functions at the clause level. The modifier functions appear later, with the moderator the last to be attested in the corpora.

5. Discussion

5.1 Two patterns of adverbialization

The way the Present-day English variation in *fairly* and other similar adverbs is described by dictionaries and grammars such as Quirk et al. (1985) requires a complex classification system. This may be taken as counterevidence to the view that *-ly* adverbialization could be accounted for as a simple inflectional process in English. Moreover, the historical evidence that we have scrutinized strongly suggests that adverbialization is a diachronic process: not all functions are available to all adverbs at all times. Some adverb categories such as intensifiers can be said to exhibit a high degree of grammaticalization in the closed-class sense of the term. Not unlike prepositions and pronouns, they contain subcategories whose membership cannot be augmented freely. These cases may be interpreted as instances of grammatical elements becoming more grammatical; as such they are not typical products of inflectional processes.

However, if we move towards the more syntactic end of the scale of fusion, the inflectional analysis seems to give an adequate description of the productivity of *-ly* adverbialization with regard to the adverbial function of adverbs, and of manner adverbs in particular. When this domain of *-ly* suffixation is considered, *adverbialization* should be taken to mean 'adverbial formation', not 'adverb formation' in general. It is synchronic in that it applies quite freely to the vast majority of adjectives at any given time. However, as this process is still gaining ground in English, a more delicate analysis would place it in the transition area between derivation and inflection. Pattern 1 describes this process of adverbialization schematically, while Pattern 2 shows the historical process of adverb formation described above.

Pattern 1: adj. 1 → adv-*ly* 1
adj. 2 → adv-*ly* 2
… …
adj. n → adv-*ly* n

Pattern 2: adj. 1 → adv-*ly* 1 → adv-*ly* 2

Pattern 1 maximizes the productivity of *-ly*, suggesting that adverbs in their typical adverbial function are formed from adjectives and continue to display regular correspondences with their adjectival bases. Pattern 2, by contrast, suggests that the link between an adverb and its adjectival base may be severed once the *-ly* adverb has been formed, i.e. that the adverb can acquire new senses

and functions independently from its base. Pattern 2 therefore predicts that some adverbs are likely to become, in Bybee's (1985) terms, more fused in the course of time. Together, the two patterns complement each other.

5.2 The sociolinguistic angle

But grammar only takes us so far, and the corpus evidence we have does not come equipped with all the information we may need to understand patterns of grammaticalization as they unfold through time. A plausible case can be made for the history of the two adverbs *fair* and *fairly* having been affected by external factors as well. First of all, we know that the suffixless form *fair* did not completely disappear. *The Oxford English Dictionary* marks most of its senses as obsolete; some of them, such as 'completely', 'fully' and quite' are also marked as dialectal.

The rise of *fairly* and disappearance of *fair* roughly coincide with the era of normative grammar. We only need to think of how Lowth and his fellow grammarians condemned the use of suffixless intensifiers to start wondering whether the development of the two adverbs might not have been interfered with by the doctrine of correctness in the eighteenth and nineteenth centuries. Pounder (2001:352) draws the general conclusion concerning adverb formation that although no "major systemic change was brought about in a dramatic way by standardization processes, the contours of the changes appear to have been moulded by them."

As shown by Biber et al. (1999:542–544) suffixless adverbs continue to be used in colloquial speech more frequently than in writing even today. The use of *-ly* adverbs may not only be more frequent in writing than in speech but it may also reveal social differences. Ronald Macaulay (1995) reports that in his study of Scottish English *-ly* adverbs were used much more frequently by his middle-class informants than by his lower-class speakers. Most strikingly, middle-class speakers used manner adverbs ten times as frequently. Moreover, in Macaulay's data, middle-class speakers used evaluative adjectives almost ten times as frequently as the lower-class speakers.

6. Conclusion

We still do not know everything about the generalization of the *-ly* suffix in social and regional terms. The evidence available suggests that it may be spreading

from above in terms of social status. Some of the more recent developments such as the application of *-ly* to colour adjectives also point to literary rather than vernacular sources. A fairly wide stretch on the cline of grammaticality would therefore probably have to be covered by anyone proposing to write a register grammar of English adverb usage in its social context.

Notes

1. This research was supported by the Academy of Finland Centre of Excellence funding for the Research Unit for Variation and Change in English at the Department of English, University of Helsinki.

2. For definitions of grammaticalization, see Campbell and Janda (2001). They conclude that "we are left with a notion of grammaticalization which minimally includes, at its core: *some linguistic element* > *some more grammatical element*" (Campbell and Janda 2001: 107). For the history of grammaticalization both before and after Meillet, see Heine et al. (1991) and Hopper and Traugott (1993: 18–31).

3. Pounder (2001: 345), too, identifies *-ly* forms with *look*, *seem* and *feel* as adverbs, but attributes their use to hypercorrection, or overapplication of adverb marking, arising from the prescriptive code.

4. The problem is not encountered to the same extent either with earlier materials such as the *York-Helsinki Parsed Corpus of Old English* or with later periods, as in the *Parsed Corpus of Early English Correspondence* (PCEEC), under construction in Helsinki and York, or the *Representative Corpus of Historical English Registers* (ARCHER). For information on the PPCME and its manual, see http://www.ling.upenn.edu/mideng/ and http://www.ling.upenn.edu/~kroch/home.html

5. However, Middle English homonymy and homography have also been disambiguated at a high level of accuracy for such more specific projects as the Early Middle English project, aimed to produce a linguistic atlas of Early Middle English (Laing 1994), and the project for producing an atlas of Older Scots (Williamson 1992/3; see also Meurman-Solin 2001).

6. But the *-ly* suffix is not fully productive either: adverbs are not usually formed from *-ly* derived adjectives such as *friendly* and *godly*. This is one argument against *-ly* adverbialization being an inflectional process (Menn and McWhinney 1984, Pounder 2001: 319, 339–340).

7. Intensifiers are morphologically perhaps the most versatile category of adverbs in English. A glance at their history would appear to support the layering hypothesis. There are intensifiers that may be called fused forms, such as the suffixless *very* and compound *somewhat*, which both go back to Late Middle English, whereas the phrasal expressions *sort of* and *kind of* are more recent.

8. The details of the various corpora referred to are given in footnote 1 in Nevalainen and Rissanen (2002). Except for ARCHER, all the historical corpora used were grammatically unannotated.

References

Aarts, B. 2004. "Modelling linguistic gradience". *Studies in Language* 28.1.

Andersen, G. 2001. Pragmatic Markers and Sociolinguistic Variation: A Relevance-theoretic Approach to the Language of Adolescents. Amsterdam and Philadelphia: John Benjamins.

Baayen, R. H. and Renouf, A. 1996. "Chronicling *The Times*: productive lexical innovations in an English newspaper". *Language* 72 (1), 69–96.

Bäcklund, U. 1973. *The Collocation of Adverbs of Degree in English* (Studia Anglistica Upsaliensia 13). Uppsala: Uppsala University.

Berglund, Y. 2000. "'You're gonna, you're not going to': A corpus-based study of colligation and collocation patterns of the *(BE) going to* construction in Present-day spoken British English". In *PALC '99: Practical Applications in Language Corpora*, B. Lewandowska-Tomaszczyk and P. J. Melia (eds.), 161–192. Frankfurt am Main: Peter Lang.

Biber, D., Conrad, S. and Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

Biber, D., Johansson, J., Leech, L., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Bolinger, D. 1961. *Generality, Gradience and the All-or-none*. 's-Gravenhage and Dordrecht: Mouton.

Brinton, L. 2002. "Grammaticalization versus lexicalization reconsidered: on the *late* use of temporal adverbs". In *English Historical Syntax and Morphology*, T. Fanego, M. J. López-Couso and J. Pérez-Guerra (eds.), 67–97. Amsterdam and Philadelphia: John Benjamins.

Bybee, J. L. 1985. *Morphology. A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Campbell, L. and Janda, R. 2001. "Introduction: conceptions of grammaticalization and their problems". *Language Sciences* 23, 93–112.

Chambers, J. K. 1995. *Sociolinguistic Theory*. Oxford: Blackwell.

Donner, M. 1991. "Adverb form in Middle English". *English Studies* 72: 1–11.

Gisborne, N. 2000. "The complementation of verbs of appearance by adverbs". In *Generative Theory and Corpus Studies; A Dialogue from 10 ICEHL*, R. Bermúdez-Otero, D. Denison, R. M. Hogg and C. B. McCully (eds.), 53–75. Berlin and New York: Mouton de Gruyter.

Givón, T. 1984. *Syntax: a Functional-typological Introduction*. Volume 1. Amsterdam: John Benjamins.

Givón, T. 1990. *Syntax: a Functional-typological Introduction*. Volume 2. Amsterdam: John Benjamins.

Halliday, M. A. K. 1961. "Categories of the theory of grammar". *Word* 17.3, 241–292.

Halliday, M. A. K., McIntosh, A. and Strevens, P. 1964. *The Linguistic Sciences and Language Teaching. London*: Longmans, Green and Co.

Heine, B., Claudi, U. and Hünnemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*. Chicago and London: Chicago University Press.

Hopper, P. J. 1991. "On some principles of grammaticalization". In *Approaches to Grammaticalization*, Vol.1, E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam and Philadelphia: John Benjamins.

Hopper, P. J. and Traugott, E. C. 1993. *Grammaticalization*. Cambridge. Cambridge University Press.

Huddleston R. and Pullum, G. K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Janda, R. D. 2001. "Beyond 'pathways' and 'unidirectionality': on the discontinuity of language transmission and the counterability of grammaticalization". *Language Sciences* 23, 265–340.

Jespersen, O. 1924. *The Philosophy of Grammar*. London: George Allen and Unwin.

Johansson, S. 1985. "Grammatical tagging and total accountability". In *Papers on Language and Literature Presented to Alvar Ellegård and Erik Frykman* (Gothenburg Studies in English 60), S. Bäckman and G. Kjellmer (eds.), 208–219. Gothenburg.

Jones, C. 1988. *Grammatical Gender in English: 950 to 1250*. London: Croom Helm.

Killie, K. 2000a. "On the use of current intuition as a bias in historical linguistics: The case of the LOOK + -ly construction in English". In *Generative Theory and Corpus Studies; A Dialogue from 10 ICEHL*, R. Bermúdez-Otero, D. Denison, R. M. Hogg and C. B. McCully (eds.), 77–101. Berlin and New York: Mouton de Gruyter.

Killie, K. 2000b. *Stative Adverbs in English: A Study of Adverbial Productivity and Orientation*. Dissertation. University of Tromsø.

Kroch, A. 2001. "Syntactic change". In *The Handbook of Contemporary Syntactic Theory*, M. Baltin and C. Collins (eds.), 699–729. Malden, MA and Oxford: Blackwell.

Labov, W. 1994. *Principles of Linguistic Change*. Volume 1: *Internal factors*. Oxford, UK and Cambridge, USA: Blackwell.

Laing, M. 1994. "The linguistic analysis of medieval vernacular texts: two projects at Edinburgh". In *Corpora Across Centuries*, M. Kytö, M. Rissanen and S. Wright (eds.), 119–141. Amsterdam and Atlanta: Rodopi.

Lehmann, C. 1995. *Thoughts on Grammaticalization*. München: Lincom Europa.

Lowth, R. 1762 [1775]). *A Short Introduction to English Grammar*. London. [A facsimile reprint (1979) of the 1775 edition. Delmar, New York: Scholars' Facsimiles and Reprints.]

Macaulay, R. K. S. 1995. "The adverbs of authority". *English World-Wide* 16, 37–60.

Marchand, H. 1969. *The Categories and Types of Present-day English Word-formation*. 2nd ed. München: C. H. Beck'sche Verlagsbuchhandlung.

Markus, M. 2000. "Normalization of the word forms in the *Ayenbite of Inwyt*". In *Placing Middle English in Context*, I. Taavitsainen, T. Nevalainen, P. Pahta and M. Rissanen (eds.), 181–197. Berlin/New York: Mouton de Gruyter.

McEnery, T. and Wilson, A. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Menn, L. and McWhinney, B. 1984. "The repeated morph constraint". *Language* 60, 519–541.

Meurman-Solin, A. 2001. "Structured text corpora in the study of language variation and change". *Literary and Linguistic Computing* 16:1, 5–27.

Nevalainen, T. 1997. "The processes of adverb derivation in Late Middle and Early Modern English". In *Grammaticalization at Work: Studies of Long-term Developments in English*, M. Rissanen, M. Kytö and K. Heikkonen (eds.), 145–190. Berlin and New York: Mouton de Gruyter.

Nevalainen, T. 2000. "Processes of supralocalisation and the rise of Standard English in the early Modern period". In *Generative Theory and Corpus Studies; A dialogue from 10*

ICEHL, R. Bermúdez-Otero, D. Denison, R.M. Hogg and C.B. McCully (eds.), 329–371. Berlin and New York: Mouton de Gruyter.

Nevalainen, T. and Raumolin-Brunberg, H. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Longman.

Nevalainen, T. and Rissanen, M. 2002. "*Fairly pretty* or *pretty fair?* On the development and grammaticalization of English downtoners". *Language Sciences* 24, 359–380.

Opdahl, L. 1997. *Ly or Zero Suffix? A Study in Variation of Dual-form Adverbs in Present-day English*. Bergen: Department of English, University of Bergen.

Palander-Collin, M. 1999. *Grammaticalization and Social Embedding; I Think and Methinks in Middle and Early Modern English* (Mémoires de la Société Néophilologique 55). Helsinki: Société Néophilologique.

Paradis, C. 1997. *Degree modifiers of adjectives in spoken British English* (Lund Studies in English 92). Lund: Lund University Press.

Pounder, A. 2001. "Adverb-marking in German and English; system and standardization". *Diachronica* 18:2, 301–358.

Quirk, R. 1965. "Descriptive statement and serial relationship". *Language* 41, 205–217.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Raumolin-Brunberg, H. 1991. *The Noun Phrase in Early Sixteenth-century English* (Mémoires de la Société Néophilologique 50). Helsinki: Société Néophilologique de Helsinki.

Raumolin-Brunberg, H. 1996. "Apparent time". In *Sociolinguistics and Language History; Studies Based on the Corpus of Early English Correspondence*, T. Nevalainen and H. Raumolin-Brunberg (eds.), 93–109. Amsterdam and Atlanta: Rodopi.

Rissanen, M. 2000. "Standardization and the language of early statutes". In *The Development of Standard English, 1300–1800: Theories, Descriptions, Conflicts*, L. Wright (ed.), 117–130. Cambridge: Cambridge University Press.

Rissanen, M., Kytö, M. and Heikkonen, K. (eds.) 1997. *Grammaticalization at Work: Studies of Long-term Developments in English*. Berlin and New York: Mouton de Gruyter.

Ross, J.R. 1972. "The category squish: Endstation Hauptwort". In *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*, P. Peranteau, J.N. Levi and G.C. Phares (eds.), 316–328. Chicago: CLS.

Sapir, E. 1921 [1978]. *Language*. London, Toronto, Sydney and New York: Granada Publishing.

Sundby, B., Bjørge, A.K. and Haugland, K.E. 1991. *A Dictionary of English Normative Grammar 1700–1800*. Amsterdam and Philadelphia: John Benjamins.

Swan, T. 1997. "From manner to subject modification: adverbialization in English". *Nordic Journal of Linguistics* 20, 179–195.

Traugott, E.C. 1982. "From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization". In *Perspectives on Historical Linguistics*, W.P. Lehmann and Y. Malkiel (eds.), 245–271. Amsterdam: John Benjamins.

Traugott, E.C. 1988. "Is internal semantic-pragmatic reconstruction possible?" In *On Language, Rhetorica, Phonologica, Syntactica: A Festschrift for Robert P. Stockwell*, C. Duncan-Rose and T. Vennemann (eds.), 128–144. London and New York: Routledge.

Traugott, E.C. and Heine, B. (eds.). 1991. *Approaches to Grammaticalization*. 2 vols. Amsterdam and Philadelphia: John Benjamins.

Williamson, K. 1992/93. "A computer-aided method for making a Linguistic Atlas of Older Scots". *Scottish Language* 11/12, 138–173.

Have to, gotta, must

Grammaticalisation, variation and specialization in English deontic modality

Sali Tagliamonte
University of Toronto

Quantitative analysis of forms used to express obligation/necessity in a variety of northern British English reveal that *must* is decreasing across generations. Instead of a marked increase in *got to* and *gotta* as reported for southern varieties of British English, these forms are used very little. There is stable variability between *have to* and *have got to*.

Multivariate analysis of internal and external factors contributing to the different forms demonstrates that *have got to* is favoured for generic statements, while *have to* is favoured for stative, personal statements. However, there is no significant effect of age or sex. It is suggested that *have to* and *have got to* both remain vigorous in this variety due to specialization of their respective functions.

Introduction

According to Bolinger (1980) the modal auxiliary system of English is undergoing "wholesale reorganization". Indeed, in a recent study, Krug (1998) observes that *have got to* for the expression of necessity and/or obligation is one of the biggest success stories in English grammar of the last century. Such claims suggest that synchronic data spanning several generations in apparent time may provide insight into the mechanisms underlying ongoing grammaticalization processes in this area of grammar. The York English Corpus, a large corpus of data from the city of York in northeast England which was collected in 1996 from speakers aged 17–95 (Tagliamonte 1998), provides an excellent opportunity to test this possibility.

These materials exhibit robust variation of a number of different forms used for the expression of necessity and/or obligation, as in (1).

(1) a. I'***ve got to*** cycle all the way back and then this afternoon I'll be cycling back up again!" You ***have to*** keep those thoughts err thoughts to yourself. (YRK/X/474,83)[1]
b. If you have a school you ***have to*** have a team and you ***have to*** have the best ones in it. But the others mustn't feel as if they're not wanted. So you'***ve got to*** steer a very close line. (YRK/§/154,47)
c. It's an old custom or something that they must- one day ***has to*** be kept for putting stalls all down. (YRK/o/267,90)

Because grammaticalization is a longitudinal process which may go on for centuries, such variability in the grammar can be interpreted to reflect the varying layers of grammaticalization attained by the different forms (Hopper 1991: 23). In order to contextualize these forms in terms of their development and history, consider the history of the modal *must* and its later quasi-modal variants *have to* and *have got to*, as outlined in Table 1.

Must has been around since Old English when its form was *mot.* Originally, it expressed permission and possibility (Warner 1993: 160–1).[2] But by the Middle English period a wider range of meanings had developed, including both deontic readings, as in (2a), and epistemic readings, as in (2b), which developed from the late 14th century (Warner 1993: 180). In contemporary English the oldest form *must* has two main functions. Epistemic modality makes judgments about the possibility, probability, or belief that something is, or is not, the case (Palmer 1979: 50). Deontic modality on the other hand (which has

Table 1. Summary of the development of deontic modality in English (Jespersen 1961: 51–54, Visser 1963–73: 1478–9, Warner 1982: 163)

Old English	**Middle English**	**19th century**	**20th century**
mot:	*Mot > must*	*must:*	*must:* "old, established"
permission or possibility	develops fuller range of senses, including *epistemic*	*have to* attested	*'ve got to* present
Primarily *deontic* modality	loses permission reading	*have got to* develops	*'ve got to*: considered "American"
	have to present, but rare		*got to* develops

also been referred to as root modality, among other terms) refers to a continuum of meanings, including obligation, permission and necessity.

This developmental trajectory of epistemic modality out of an original deontic modality is a common path in grammatical change, both in English and cross-linguistically (Bybee et al. 1994, Krug 1998, Traugott 1997).

By the end of the Middle English period *must* replaced *mot*[3] and the permission reading of *must* was lost (Denison 1993:303).

(2) a. De ***mote*** neh gon; & neodliche heom fon on …
'You *must* go hear, and forcibly take hold of them …' (1200?)
b. … goodis lawe ***mut*** passe in authorite mannuys law …
'… God's law *must* surpass man's law in authority.' (1378?)
(examples from Warner 1993:175)

According to the *Oxford English Dictionary* (OED) the use of *have to* in the sense of "obligation" is first attested in 1579, as in (3).

(3) He told him, he ***had not to*** beleeue, that the couetousnesses of Virginio .. had moued Ferdinand. (Fenton Guicciard)

However, Crowell (1955) suggests it may have been even earlier, providing the example in (4) from the Paston Letters.[4]

(4) Every man … ***hath*** greatly ***to*** marvel that any man would say the loss of two so noble duchies … is but trespass. (Paston Letters, No 191 I. 260)

Layering in the system was apparently present from at least that time onwards, as observable in (5) where *must* and *have to* alternate in a single sentence.

(5) I ***moot*** go thider as I ***haue to*** go. (Chaucer, CT. Pard. C. 749)
'I must go thither as I have to go.' (Brinton 1991:34)

The construction *have got to* on the other hand, as in (6a), or with *got* by itself, as in (6b), entered the English language much later — not until the 19th century (OED s.v. *get*, no 24 (Visser 1963–73:479). Both Visser and the OED label it as colloquial, even vulgar. In fact, prescriptive grammars have long regarded it as somewhat stigmatized and present-day English grammars usually consider it "informal".

(6) a. He always remembers when I'***ve got to*** take my doctor's stuff. (1860, George Eliot, Mill on the Floss I, IX.)
b. … they ***got to*** stop worry about about themselves. (1942, Joyce Cary. To Be a Pilgrim, 65) (examples from Visser 1963–73)

However, in a recent large-scale analysis of the British National Corpus of English, Krug (1998) demonstrated that referring to *have got to* or *gotta* as simply "informal" is quite an understatement. He found that in British English of the 1990's *have got to* and *gotta* were one and half times as frequent as the older forms *must* or *have to.*

According to this general trajectory, it would seem that the construction with *got* is grammaticalizing and further that it is taking over as the marker of deontic modality in English.

This scenario of long-term evolution of forms for the same function, yet recent and rapid takeover by one construction in particular, presents an interesting subject for a study of grammaticalization processes in corpus data. First, because the forms entered the language at very different points in time, their distribution can shed light on the stages of development of the system in a given corpus and thus the nature of linguistic change (particularly grammaticalization) in this area of English grammar (Bybee et al. 1994). Second, comparisons with patterns extrapolated from the historical and synchronic literature can be used to track varying trends in corpus data and illuminate the underlying mechanism involved in the grammaticalization process.

In this paper, I shall demonstrate that a representative community-based sample of spoken vernacular language data provides an important picture of varying stages in the grammaticalization process. Moreover, the information that can be gleaned from non-standard vernacular data sheds light not only on community-based, regional norms, but also situates linguistic change at a particular point on its trajectory. Second, I explore the utility of grammaticalization theory for the study of morpho-syntactic developments in English where linguistic variation is indicative of ongoing change. Internal linguistic constraints (or correlations) on variable forms can be traced to constraints attested in the history of the English language, and thus can be interpreted as "persistence" (Hopper and Traugott 1993). This provides some insights into what earlier points in the trajectory of development of these areas of grammar may have been like. Similarly, as forms take on new grammatical functions we may observe shifts and re-weighting of contextual effects pointing to "specialization" (Hopper 1991). Indeed, differences in inter-variety distributions across generations may reveal the pathway of such change. As we shall see, the findings I report here suggest that a community represents its own "slice in time", reflected, not only in the varying frequency of forms, but more strikingly in their patterns of use (distribution). As I will argue, the critical contribution of synchronic dialect data is to illuminate ongoing processes of change. Third, I

Table 2. The York Corpus. Sample criterion: born, raised, lived in York all their lives

	Male	**Female**	**Total**
< 30	11	11	22
31–69	17	24	41
> 70	12	17	29
Total	40	52	92

demonstrate the utility of variationist sociolinguistic methods in the analysis and interpretation of linguistic patterns and the critical role it serves in their evaluation (Poplack and Tagliamonte 2001:88–12). In order to assess the grammatical function(s) of forms and their status in the community, I test the effects of linguistic features associated with the linguistic change. I then correlate these contextual factors with the different variants in the data to assess their patterning using quantitative techniques such as distributional analysis. Further analysis using multiple regression can assess the direction of effect of these patterns, their significance and relative importance when all factors are considered simultaneously. The comparative method is then used to assess similarities and differences across age groups in the community. For further discussion see (Poplack and Tagliamonte 2001: Chapter 5, Tagliamonte 2002).

The data come from York, a small city in northeast England, as represented by a cross-section of the population stratified by age, sex and socio-economic class as shown in Table 2. The speakers who range in age from 17–92 were both born and raised in the city. Thus, this data should provide a viable model of this community's variety of English.

Method

Like many morpho-syntactic features deontic modality is quite rare. The entire 1.2 million word York English corpus was exhaustively searched for tokens of *must*, *have to*, *have got to*, *got to*, and *gotta*, and any of their phonological variants. However, raw frequencies of forms do not take into account functional differences amongst them. Therefore, it was crucial to establish which contexts were actually part of the relevant set of 'layers'. Following Coates (1983:32), I restrict the analysis to all contexts in which *must*, *have to*, or *have got to* was possible and which also encoded the meaning 'it is necessary for …'.

Contexts involving past or future markers, as in (7) were excluded. Consistent

with descriptions of the English modal system these were categorically used with *have to*.[5] *Have (got) to* had no non-finite forms (Palmer 1979: 114).

(7) a. I was so glad that I ***didn't have to*** have anything amputated. (YRK/049/80)
b. I think I'***ll have to*** do some shopping this afternoon. (YRK/039/35)
c. I ***used to have to*** get club cheques out to clothe the lads. (YRK/003/67)

At first, all uses of *must* were extracted, regardless of meaning, in order to make an independent assessment of the forms used for different semantic functions of *must*.[6]

Use of *must* for epistemic modality, as in (8) and (9), was very frequent. In fact, the contexts in (8) were categorically rendered with *must* while those in (9) were near categorical.

(8) a. It ***must have been*** when York races were on. (YRK/068/14)
b. In my teens in York there ***must've been*** nine cinemas. (YRK/001/11.5)

(9) a. It ***must be*** something to do with kangaroos or something. (YRK/"/315,35)
b. Our Paul's twenty oh, oh good grief, he ***must*** be twenty-four. (YRK/s/488,63)
c. You ***must*** know where that is. (YRK/t/16,9)
d. … very good dinner once a year which ***must*** cost quite a lot of money. (YRK/E/981,35)
e. They ***must*** have really good parents, that's all I can say. (YRK/'/357,17)

In addition, *must* is often used in formulaic expressions, as in (10):

(10) a. 'Cos they say "Oh it's less painful." I was like "You ***must*** be joking. (YRK/W/407,92)
b. I feel really bitter about it, I ***must*** admit (YRK/R/411,25)
c. I ***must*** confess, I've known Maurice er, nearly forty years. (YRK/=/222,28)

Table 3 shows the distribution of these different semantic types of *must*. It reveals that epistemic readings are lexicalized by *must* over 80% of the time with only rare occurrences of *have to, have got to* or *got to*.

This distribution pattern corresponds to that found by Coates (1983: 48)

Table 3. Overall distribution of variants of epistemic modality

must		*have to*		*'ve/'s got to*		*got to/gotta*	
%	*n*	%	*n*	%	*n*	%	*n*
81	63	9	7	8	6	2	2

and Palmer (1979: 53) for contemporary (British) English (Survey of English Usage) — epistemic *must* is holding is own, but deontic *must* is receding dramatically.

In contemporary English dialects *have to* and *got to/gotta* may also be used for these epistemic readings in addition to *must*, as you can see in (11). However, these uses are generally considered a later, and particularly American, development (Coates 1983: 57). In the York data they are extremely rare. There were only 15 in the entire corpus.

(11) a. There ***has to*** be something more than this. Is there somewhere that we're missing? (YRK/d/380,50)
b. I think there must be and they'***ve got to be*** a lot more intelligent than us, haven't they? (YRK/¡/884,50)

This suggests that British English, and/or York English in particular is lagging behind in ongoing linguistic change in this area of grammar. Moreover, the dramatic difference in the distribution of *must* for each context reveals how important it is to take function into account when examining grammaticalization processes. Had the frequency of the forms been tallied without consideration of function (i.e. epistemic vs. deontic reading), the divergent usage patterns in each function domain would have been obscured. In the end I was left with a total of 461 tokens of deontic modality from 92 speakers and approximately 110 hours of running conversation.

Another factor to consider in the analysis of the forms used for deontic modality in English is their syntactic properties as these provide important insight into the categorial status of the constructions in addition to their stage of grammaticalization.

True modals are operators. As such, they can be negated, take part in subject–verb inversion, survive post-verbal ellipsis and be stressed. However, only *must* is considered a true modal. *Have to* is generally thought to be a quasi-modal. On the one hand, it behaves like a modal because of its meaning. But syntactically it behaves like a main verb (Denison 1993: 316). Auxiliary-like uses of *have to* can be observed in contexts in which it is negated, inverted, ellipted,

Table 4. Overall distribution of variants in negative constructions

don't have to		*mustn't*		*must not*		*haven't got*	
%	*n*	%	*n*	%	*n*	%	*n*
79	11	7	1	7	1	7	1

or stressed, the standard tests for auxiliary status.

The most frequent of such contexts in the data were negative constructions.[7] Examination of the distribution of forms in these contexts, as in (13), reveals first of all that negation is extremely rare — there were only 14 in all, as in Table 4.

Auxiliary-like uses of *have to* in these contexts, as in (12), started developing around the late 19th century (Denison 1993: 317, Visser 1963–73: §1467). However, Warner (1993: 12) points out that such uses were rare. These data corroborate this. There is not a single instance in this corpus.[8]

(12) You ***haven't*** to say anything. (Denison 1993)

Instead 79% of them are negated with *do*-support and *have to*, as in (13b) and (13c).[9] Thus, with respect to negation *have to* is still functioning like a main verb.

(13) a. One thing I ***haven't got to*** do is …. (YRK/H/468,0)
b. No grief off anybody. It's your day, ***don't have to*** please anyone. (YRK/¡/119,74)
c. I ***don't have to*** give you a reason. (YRK/@/168,69)

Denison (1993: 317) observes that although *have to* does have auxiliary-like properties in northern British dialects, it is either receding or was never there in the first place.[10] More recently, it is *got to* that appears to be showing evidence of a change in categorial status towards more auxiliary-like behaviour. For example, Krug's (1998) study revealed that it can be used in ellipsis and is starting to take *do*-support. Again, however, there were no uses like this in these data either.[11]

Thus, variability (or layering) of deontic modality in York is actually quite circumscribed. The vast majority of contexts are affirmative contexts of obligation/necessity. Moreover, this is where there is layering. Within these contexts, however, there is little evidence for a change in the categorial status of the older variant, *have to*, nor development of the new variant *gotta*, suggesting a conservative stage for deontic modality. Of course, in the process of change linguistic forms only gradually shift from one function to another. Critical for

my purposes is that this trajectory can also be viewed in the varying strength and distribution of independent linguistic features associated with one of the evolving grammatical morphemes. Indeed, such environmental correlations are held to be the keys to viewing the mechanism of diachronic grammaticalization in synchronic data (Traugott 1991). I now turn to an analysis of the variable contexts of deontic modality according to their distribution and contextual linguistic correlates.

Analysis

As mentioned earlier, each of the contexts retained for analysis was coded for a series of factors implicated in the development of deontic modality. Table 5 presents the overall distribution of forms. The frequency of *got to/gotta* is very low, only 3% of the data. This is consistent with the historical trajectory which shows that *gotta* is a late development in this area of grammar. *Must* is much more frequent, occurring 15% of the time. However, the most interesting result is the extent of variation between *have to* and *have got to*. The rates of use are identical.

Table 5. Overall distribution of variants of root modality

must		*have to*		*'ve/'s got to*		*got to/gotta*	
%	*n*	%	*n*	%	*n*	%	*n*
15	67	41	187	41	189	3	15

However, recall that this overall distribution combines all the speakers in the community. Because this is an area of grammar known to be undergoing change we might expect to find a correlation between the incoming variant and speaker age.

Figure 1 provides an overall distribution of forms, however this time separating the data into three broad age groupings, over 70 years, 31–69 years, and under 30 years. First, consider the oldest variant in this system — *must*. This form is thought to be declining rapidly, and considered to be old and somewhat archaic. Here, we see a graphic confirmation of this. The greatest frequency of use is in the speakers over 70, with a smaller proportion amongst the middle-aged generation. The under 30 year olds, on the other hand, hardly

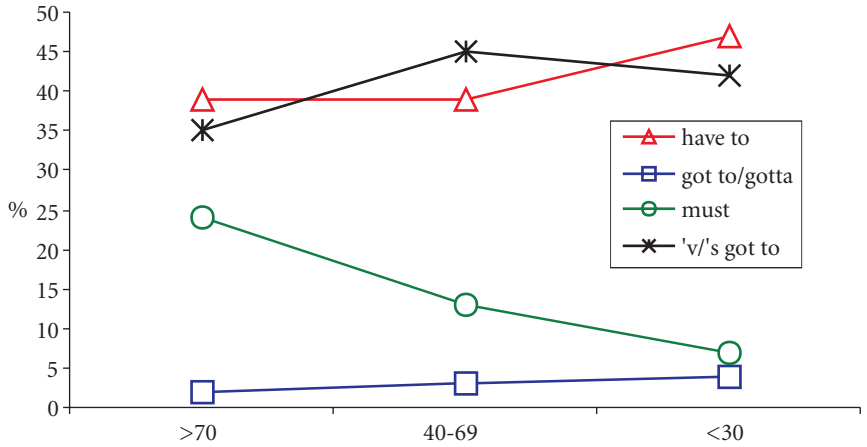


Figure 1. Overall distribution of variants of DEONTIC MODALITY (obligation/necessity).

ever use it. *Must* clearly exhibits the trajectory of an obsolescing feature. Consider the next layer of development, the construction with *have to.* This form is used very frequently, and appears to be increasing in use amongst the younger generation where it is the most common form used to express deontic modality. The construction with *have got to* however is nearly as frequent. Finally, the most recent layer in the development of deontic modality — *got to* or *gotta* with no corresponding auxiliary.[12] This form is hardly ever used and is restricted to the middle and younger age groups.

The extent to which deontic modality is expressed by variable use of *have to* and *have got to* across the board in this community is noteworthy. Table 5 showed that these two forms compete in this area of the grammar. Figure 1 now reveals this to be consistent for all age groups. Nothing in the literature would lead one to expect such robust and stable variability between these two constructions. Taken together, these results suggests that *have got to* is not entirely taking over as the central modal of obligation, at least not in this variety of northern British English. Indeed, Krug (1998) reported quite a dramatic split between southern and northern British dialects with respect to the use of *have got to.* Thus, the frequency of use of *have to* and *have got to* in York is a demonstration — on morpho-syntactic grounds — of one of the most important dialect boundaries in Britain: that between north and south. As Trudgill (1990: 65–78) and others have argued, northern British dialects are generally more conservative than the southern ones. In this case, the robust use of *have*

to in York is consistent with the fact that northern dialects in general are relatively conservative. In this case we see retention of an earlier layer in the grammaticalizing system of deontic modality. The question remains however, what are these two forms *have to* and *have got to* doing? Is this a case of subtle functional specialization, sociolinguistic variation, or some other factor?

Grammaticality judgements elicited from York natives about these differences, e.g. *I have to go shopping* vs. *I've got to go shopping*, confirm that there is little to choose between the two. Thus, whatever differentiates them is beneath the level of consciousness in this community. This is consistent with Palmer's (1979:93) suggestion that "there may be complete overlap in the area of neutral necessity".

According to Coates (1983:32) root modality, of which deontic modality is a part, is "a fuzzy set" with a cline which "extends from strong to weak obligation", including at one end readings which translate as 'it is imperative that …' to those which simply mean 'it is important that …' Although subjective in interpretation, this cline of strength can be related to a number of independent grammatical diagnostics, including the animacy of the subject, the type of verb and grammatical person. Although Coates was describing a cline which was consistently rendered with *must*, the extreme variability observed in the data between *have to* and *got to* makes it plausible that these two variants may be conditioned by the same factors.

In other words, it may be that internal linguistic features such as animacy, type of verb and grammatical person are governing the choice between *have to* and *have got to*, particularly in this community where they are both used so frequently for the same function. Further, are these effects the same or different across generations?

In order to test for such effects, I shall now examine the distribution of forms and correlate them with these factors and compare the distributional patterns across generations. For clarity, I shall present all the results in terms of the proportion represented by the older variant *have to.*

Animacy

First, consider animacy. As it turns out, 97% of these data have animate subjects, as in (14).

(14) a. The following day I'***ve got to*** be at Church flower arranging for harvest festival. (YRK/x/517,71)

b. And Sundays is our relaxing evening 'cos John, unfortunately, ***has to*** work Sundays. (YRK/'/409,71)

Thus, while Coates (1983:53) suggests that *have got to* is typically associated with animate subjects, these data reveal that there is little other than animate subjects for any of the variants. Moreover, the few subjects that are inanimate (n=16), as in (15), show no preference for any one form.

(15) a. It all ***has to*** be heated and cleaned. (YRK/z/125,18)
b. My car'***s got to*** go to the garage. (YRK/a/771,17)

Animacy of the subject is relevant to the expression of deontic modality insofar as deontic constructions are overwhelmingly animate. However, there is no particular form that favours the rare inanimates that occur here. Historical research has shown that the *have to* construction started admitting inanimate subjects as early as Middle English and gradually extended into this area (Brinton 1991:36). However, there is no evidence for such expansion in this variety. This result is consistent with my earlier interpretation of this dialect as conservative.

Type of verb

Coates (1983:37) also observes that stative verbs tend to cluster more at the weak extreme of necessity and/or obligation readings, while activity, or punctual verbs, are favoured for strong readings. This would suggest that verb type may differentiate the use of *have to* as opposed to *have got to.* Figure 2 tests this correlation on the data.

The figure shows that *have to* is consistently more frequent with stative verbs, as in (16), and durative verbs, as in (17), as opposed to punctual verbs, as in (18).

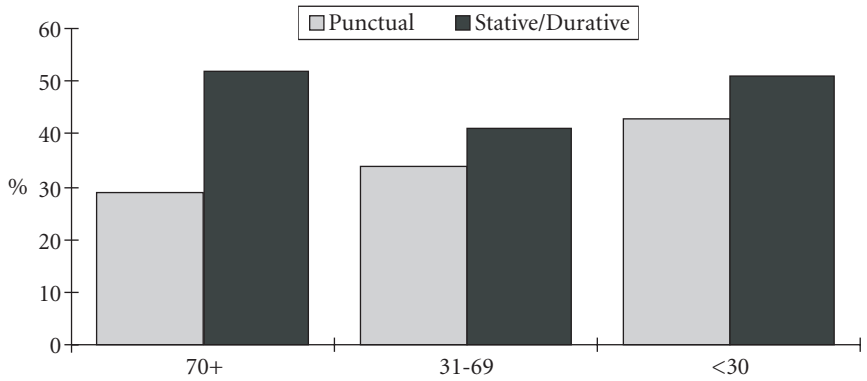


Figure 2. Distribution of *have to* across communities by type of verb.

(16) a. …just ***have to*** know what type of ball to use. (YRK/™/171,5)
b. You ***have to*** believe in ghosts to ever see it. (YRK/%/357,4)
c. But it's going to the loo that I think'***s got to*** be the worst. (YRK/n/366,65)

(17) a. You ***have to*** live with someone to know them. (YRK/c/366,102)
b. We ***have to*** try and fit it on that schedule somehow. (YRK/R/501,75)

(18) a. You'***ve got to*** put your ten pennies in. (YRK/h/468,56)
b. No, if they sack us ***they've got*** to find us a new job, like. (YRK/+/1104,29)

Thus, not only do varying verb types differentiate forms in this area of grammar. Taken with Coates' observation, *have to* can be considered to encode weak obligation.

Grammatical person

Coates (1983:37) also suggests that 2nd person subjects encode the strongest readings of necessity and/or obligation. If *must* is the stronger form, this predicts that *must* may be the over-riding choice for these subjects. Note, however, that Coates is undoubtedly referring to 2nd person subjects with definite readings, as in (19) rather than generic readings as in (20).

(19) a. The people that were here before me said "You ***don't have to*** worry," they said. (YRK/Z/99,7)
b. You ***have to*** remember that very few houses in those days had baths. (YRK/M/328,87)

(20) a. Sometimes she just can't be arsed can she? And she just sits there and meows and you'***ve got to*** open t' door for her. (YRK/&/592,90)
b. I think you can't stagnate, you'***ve got to*** keep progressing. (YRK/6/211,80)

The perspective provided by the quantitative approach to the data employed here reveals that in these data this difference is extremely important. There are widely divergent frequencies of these subject types in the data. Generic *you* makes up the vast majority (85%) of the 2nd person subjects in the data (182/215). Further, generic *you* represents a full 34% of all the grammatical subjects in the data (182/534).

Third person subjects also have both definite readings, as in (21) and generic readings, as in (22). Like 2nd person, here too generic readings are in

the majority (58%, 50/86) although 3rd person subjects do not represent nearly as high a proportion of the data as 2nd person subjects (only 16%, 86/534).[13]

(21) a. They ***have to*** hand their homework in a cupboard on the morning. (YRK/T/476,10)
b. Oh dear! Why did they ***have to*** see me in my underpants? (YRK/D?64,68)

(22) a. But when people are desperate they ***have to*** pay it, don't they? (YRK/+/1059,66)
b. Big sisters are lovely. Well, they ***have to*** be, don't they. (YRK/S/136,91)

Taken together these distributional facts reveal that any effect of 2nd or 3rd person will also be overwhelmingly biased toward generic readings. Moreover, given the high proportion of generic subjects that are 2nd person sg. *you*, any correlation by subject type will also be biased towards these contexts in particular. In other words, most of the data are examples such as in (20). On the other hand, all first person subjects are definite by definition and these subjects represent the vast majority of the non-generic subjects. It is important to keep this in mind in interpreting these data and comparing them to the predictions from the literature.

Thus, for example, the distribution of *must, have to and have got to* differs markedly when 2nd person generic subjects are separated from those that are 2nd person non-generic, as in Table 6. 2nd p. sg. definite subjects pattern with 1st p. sg. in having a fair proportion of use of *must.* On the other hand, 2nd p. sg. generic subjects have an overwhelming proportion of *have got to* (57%).

Note however that the vast majority of forms in the data are either 2nd p. generic or 1st p. singular. In order to test whether type of reference correlates with the use of *have to* and *have got to*, Figure 3 shows the distribution of these

Table 6. Distribution of variants of deontic modality by 1st p. sg. and 2nd p. sg.

	must		*have to*		*'ve/'s got to*		*got to/gotta*		Total
	%	*n*	%	*n*	%	*n*	%	*n*	*n*
1st p. sg. *I*	**29**	45	37	57	32	50	2	3	155
2nd p. sg. *you* DEFINITE	**26**	10	31	12	36	14	8	3	39
2nd p. sg. *you* GENERIC	3	8	35	80	**57**	132	4	10	230

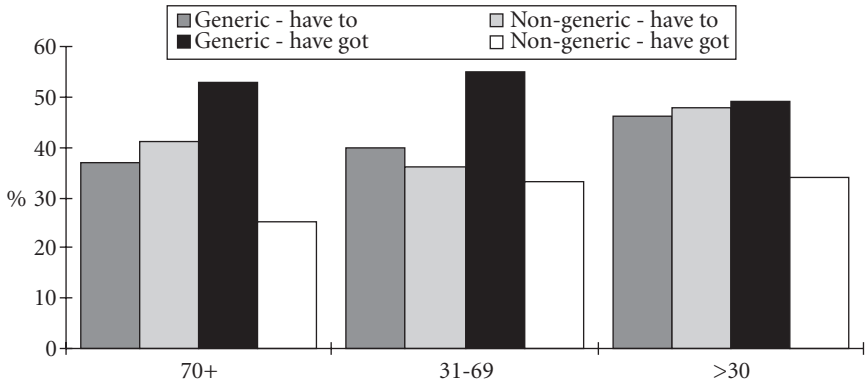


Figure 3. Distribution of *have to* and *have got to* across communities by type of reference.

forms according to whether the reading is generic or non-generic.

When the distribution of forms is compared across generic (primarily *you* generic) and non-generic (primarily first person singular, *I*) subjects, it becomes clear that there is little to differentiate the use of *have to* by generic or non-generic. As with the distribution by animacy, this is consistent across generations. On the other hand, use of *have got to* shows a distinct preference for generic subjects amongst the oldest speakers (70+) and amongst the middle-aged group (31–69). Notice however, that amongst the youngest speakers use of *have to* rises in frequency to rival the rate of *have got to* for generic subjects. Moreover, amongst this age group *have to* is just as likely to be used for generic subjects as it is for non-generic subjects.

So far, the analysis reveals that there is tremendous stability in the overall frequency of *have to* and *have got to* across generations (Figure 3). However, when we examine the correlation of these forms with two contextual factors (constraints) we find that one, (verb type) remains stable, while another (type of reference) shifts across the community. The change in the operation of these contextual constraints in apparent time hints at ongoing grammaticalization processes in the system of deontic modality more generally. However, we still do not know which of these factors, if any, are statistically significant, a finding which would support the interpretation of this change as relevant to the grammaticalization process.

Multivariate analysis which can model complex constraints and relative weights of numerous factors that operate simultaneously on linguistic features (Labov 1982: 51) provides an invaluable tool for actually tracking a

Table 7. Three independent variable rule analyses of the contribution of internal and external factors to the probability of *must*, *have to*, and *have got to* in York English. Factor groups selected as significant in bold.

	must		*have to*		*'ve got to*		
Input:	.07		.41		.41		
Tot $n = 461$	67		187		189		
Factors:	factor weight	%	factor weight	%	factor weight	%	Ns/Cell
Age							
> 70	.64	25	[.49]	39	[.46]	34	122
31–69	.54	13	[.46]	38	[.53]	45	213
< 30	.31	7	[.58]	47	[.59]	41	126
Range	*33*						
Type of verb							
Punctual	.71	25	.44	35	[.47]	36	225
Stative, Durative	.30	5	.56	46	[.53]	47	236
Range	*41*		*12*				
Type of reference							
Generic	.20	2	[.51]	41	.63	53	179
Non-generic	.71	26	[.50]	41	.42	30	282
Range	*51*				*21*		
Factors not selected in any analysis:							
Sex	X		X		X		
Education	X		X		X		

grammaticalizing linguistic feature. Like the tracks of linguistic change more generally, grammaticalization should be visible in an ordered series of shifts in factor weights (Labov 1982: 76). Table 7 displays the result of this type of analysis in these data, providing a multi-dimensional view of deontic modality in this community. The input indicates the overall tendency of the form (in this case *have to* as opposed to *have got to*) to surface in the data. Point-form numbers are factor weights, indicating the probability of the form occurring in the context indicated. Statistically significant correlations are indicated in bold face while those that are not significant are enclosed in square brackets.[14] The closer these numbers are to 1, the more highly favouring the effect is; the closer they are to zero the more disfavouring the effect is. The range indicates the

relative strength of the effect. The higher the numbers, the stronger the effect. All these calculations are computed by the step-wise selection procedure incorporated in the variable rule program (see Rand and Sankoff 1990).

The multivariate analysis provides statistical confirmation of a number of important trends and correlations within the system of deontic modality in this community. First, the rapid obsolescence of *must* in apparent time is visible within the factor of age. This factor is statistically significant and the factor weights for the probability of occurrence of *must* decrease from the oldest speakers where it is favoured at .64, to the youngest speakers where it is disfavoured at .31. Note the conditioning effects on *must* for both type of verb and type of reference. The strength of these correlations is an indication of how circumscribed this form is. It is highly favoured with punctual verbs (.71) and non-generic readings (.71). These effects pinpoint the use of *must* to clearly defined contexts in discourse. Further, as Coates (1983) suggests, these are often directive as well as performative, where the speaker has authority over the subject, as in (23). These examples involve contexts in which the speaker has authority over the subject, for example, officer to private, (23a), parent to child, (23b).

(23) a. And they just took us down, round the dock area and an officer says, "Right, you're in here." Door had about four big bolts on it. Says, "Lock every bolt." He says, "If you've anybody, tries to break- don't let anybody in on any pretense. It didn't matter who they say they are. You ***must*** not open that door until we come tomorrow morning." (YRKj/37,61)

b. Well, I have been a good girl, haven't I?" And I say "You ***must*** take these sweeties for being a good girl. (YRK/I/396,61)

However, *must* is very rare in the data; notice that the input for *all* deontic uses of *must* is only .07. Given that the factor weights vary between 0 and 1, this means that its overall propensity of occurrence is very low. Moreover, when we compare the grammatical factors which constrain the use of *have to* and *have got to* with those which constrain *must* they are quite different. *Must* is favoured for punctual verbs, but neither *have to* nor *have got to* is. *Must* is favoured for non-generic subjects, while *have to* is neutral in this regard and *have got to* is favoured in the opposite context, generics. There is no continuity in the internal constraints.

Thus, there is little evidence to suggest that the process involved here is a simple substitution of *have to* for *must* as suggested by Conradie (1987: 179). This is probably because neither *have to* nor *have got to* explicitly invokes the speaker's authority over others, something which *must* encoded explicitly.

Instead, the obsolescence of *must* may actually be tied to the obsolescence of the appropriate social conditions for its use, particularly in these type of interview data. It could also be that the appropriate contextual use for *must* has receded into a highly specialized set of circumstances that are stylistically marked. This too is part of what happens as forms grammaticalize, some forms get left behind, entrenched in formulaic utterances or certain discourse rituals. Indeed, some of these were shown earlier in (10).

In any case, the productive forms in these data are *have to* and *have got to*. How these are used across the range of contexts of necessity/obligation will provide critical evidence for the grammaticalization process. Indeed, the multivariate analysis results reveal that at the community level each one has its own particular niche in the deontic modality system. Overall, *have to* is favoured for stative and durative contexts at .56; while *have got to* is favoured for generic reference at .63. The favouring effect of stative/durative contexts for *have to* shows persistence of the original lexical meaning of the verb *have*. The specialization of *have got to* for generic readings confirms the reports in the literature that *have got to* encodes readings of strong obligation associated with 2nd person singular. However, these are mostly 2nd person generic readings. Moreover, recall that the specialization of *have got to* for generic readings reported in the literature was visible only amongst the older generations in Figure 3. Amongst the younger group the distinction between generic and non-generic appeared to be levelling out. This suggests that the overall view presented in Table 7 here may be obscuring correlations that are changing over time. In other words, in order to make visible the ordered series of shifts in factor weights, referred to by Labov (1982: 76), it is necessary to further divide the data by generation.

I now re-analyze the same data, but analyze the effects of internal correlations separately for each age group. Table 8 displays the results of this analysis.

The variable rule analysis shows that the effect of the type of verb is significant in the oldest generation; stative/durative verbs favour *have to* at .61 while punctual verbs disfavour at .41. In the middle generation the ranking of stative/durative verbs is still visible (.53 > .46), but the factor does not achieve statistical significance. Finally, in the youngest generation the constraint is not statistically significant, nor is there any ranking of one context over the other (.50 = .50). These results provide statistical confirmation of the shifting weight of the type of verb constraint in apparent time. The effect of generic on *have to*, however, is stable across the board indicating that this factor is not implicated in the use of *have to*.

Table 8. Three independent variable rule analyses of the contribution of internal and external factors to the probability of *have to* in York English by age group. Factor groups selected as significant in bold.

	Old		**Middle**		**Young**	
Input:	.37		.40		.41	
Tot $n = 461$	124		210		127	
Factors:	factor weight	%	factor weight	%	factor weight	%
Type of verb						
Stative, Durative	**.61**	49	[.53]	41	[.50]	47
Punctual	**.41**	29	[.46]	34	[.50]	43
Range	*20*		*12*			
Type of reference						
Generic	[.46]	34	[.51]	40	[.50]	40
Non-generic	[.53]	40	[.49]	43	[.50]	40

Discussion

This paper has analyzed the system of deontic modality in English using a dialect corpus as data and grammaticalization theory as a means to interpret the findings. This approach has revealed a *layering* of forms which encode the meaning of obligation/necessity and provided a detailed characterization of this system in a variety of northern British English. The qualitative nature of the system is outlined in (A):

A. Characterictics of deontic modality in York English
   - Epistemic *must* is far more frequent than deontic *must*.
   - Epistemic uses of *have to* or *have got to* are rare.
   - Contexts of deontic modality are mostly affirmative, animate, active verbs, very often encoded with 2nd person (generic) subjects.
   - Use of *must* is highly circumscribed to performative, directive contexts in which the speaker has authority over the subject.

In the context of a changing grammatical sub-system, an analysis of forms across the different age groups in the community (apparent time) revealed a trajectory of change. This may be interpreted as a "slice in time" in the ongoing grammaticalization of the system of deontic modality, whose characteristics are as in (B).

(B) Conservatism?
- *Must* is present, but decreases dramatically in apparent time.
- Use of *have to* is robust across all generations of speakers in the community.
- There is little or no extension to inanimate subjects.
- The use of *got to* or *gotta* is rare.
- There is no modal-like uses of *got to/gotta.*

The nature of these characteristics contributes to a plausible assessment of the variety under investigation York English — as "conservative" with regard to deontic modality. They differentiate the York data from what is reported for other varieties of English, in particular British dialects in the south (e.g. Krug 1998) but also mainstream North American dialects. In both the latter locales the new variant *gotta* is said to have made considerable progress in supplanting the other forms.

The York English data also exhibits correlations for the forms used for obligation/necessity which are consistent with reports in the literature, e.g. the *specialization* of *have got to* for generic readings. When the data are analyzed overall (Table 7) the specialization of *have to* for stative/durative readings and *have got to* for generic readings is confirmed — in both cases the correlation is statistically significant. However, there is also evidence for ongoing re-weighting of internal grammatical constraints. When the correlations are tested separately by speaker age, there is a hint of ongoing change (Figure 3). Table 8 confirmed that the correlation of *have got to* for generic readings in the older generations is levelling out in the younger generations. Lapses in co-occurrence restrictions, here affecting the type of subject or the type of verb collocated with *have to* and *have got to*, may be expected to occur as an item is generalizing in meaning. Indeed, as forms grammaticalize, there is predicted to be a diminution of formal choices and the surviving forms will assume more general meanings. This is what we observe as the original meaning of *have* (correlated with certain types of verbs) is replaced with a more general reading of necessity/obligation. In this way, the extension of use of *have to* according to type of subject and type of reference can be interpreted as the result of ongoing grammaticalization.

In sum, I suggest that we can interpret the ebb and flow of change in this system of grammar tentatively as follows: as *must* obsolesces, it remains entrenched for readings with strong obligation (i.e. specific subject reference with action verbs (e.g. You *must* get something in your own name. [YRK/x/445, 54])). However, where once *have to* and *have got to* were somewhat specialized

elsewhere within this system (*have got to* for generic reference and *have to* for stative/durative verbs), the differences between them are diminishing.

Taken together, the results of this single community-based study can now be used as a base-line to compare and contrast from a wider perspective. Do the same frequencies of *must, have to* and *have got to* exist elsewhere? Does the layering between *must, have to* and *have got to* pattern the same way elsewhere? Is the trajectory of development similar or different? Are the developments in apparent time mirrored in real time? Broader study of the system of deontic modality, both across dialects in the UK, and across major varieties of English, will present a unique opportunity to track the developments of a sub-system of grammar at a time when different present-day varieties preserve varying stages in its development. Indeed, this area of grammar presents a unique opportunity to tap into a linguistic change as it is occuring (see Labov 1972: 323). It will be interesting to find out how the competition between *have to* and *have got to* is resolved if *must* continues to fade away.

Notes

1. The information in parentheses contains first, the codes for York (YRK), second, the speaker identification symbol and 3rd the co-ordinates of the datum in the data files.

2. In late Old English it developed the modal meaning of subjective deontic necessity, which is the meaning that some authority grants permission for an action.

3. Some scholars argue that the present day English modal semantic system was established by Shakespeare's time (Ehram 1966: 97). Others argue that there has been "a continued semantic focussing of modality" in which "the modals continue to lose past-referring uses of their preterite forms and there is continued reduction in the relevance of subject-oriented uses" (Warner 1993: 181).

4. Van der Gaff (1931), cited in Denison (1993: 316), thinks that the "obligation" reading for *have to* was present in Old English. Mitchell argues against this (1985: 950–3).

5. See Palmer (1979: 114) "*have got to* has no non-finite forms … *have to* must be used".

6. These were categorized separately for comparison and excluded from the main analyses.

7. There were only two question forms and no stressed contexts.

8. In contrast to use of *have* for possessive meaning, which is used far more frequently in the north than elsewhere (see Tagliamonte 2003).

9. Consistent with descriptions in the literature, the negative forms of *have to* and *got to* negate the modal predicate in contrast to *must* where it is the main predication which is negated.

10. Denison (1993: 317) also suggests that *have to* has quite recently developed auxiliary-like properties in the southern standard, e.g. *You haven't to say anything.*

11. There is also anecdotal evidence of this: Trudgill reports that some British dialects use *don't gotta* productively. However, given the late onset of these developments, the expectation for the dialects under investigation would be that such constructions would be negligible or non-existent and they are. Nevertheless, Trudgill's observations point to areas for further research elsewhere.

12. The two surface forms *got to* and *gotta* were coded separately; however, there were only two instances of *gotta*. These were collapsed with the infrequent ($n = 15$) *got to* tokens due to the small numbers.

13. Use of generic *you* increases slightly from oldest to middle-aged to youngest speakers, i.e. 30% < 43% < 40%.

14. Statistical significance is calculated at the .05 level.

References

Bolinger, D. 1980. *Language. The Loaded Weapon.* London: Longman.

Brinton, L. 1991. "The origin and development of quasimodal *have to* in English". Paper presented at the Workshop on Verbal Periphrases. Amsterdam. August 1991.

Bybee, J. L., Perkins, R. D. and Pagliuca, W. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Coates, J. 1983. *The Semantics of the Modal Auxiliaries.* London: Croom Helm.

Crowell, T. L. 1955. "Predating 'have to', 'must'". *American Speech* 30(1): 68–69.

Denison, D. 1993. *English Historical Syntax.* Harlow, Essex: Longman.

Ehram, M. E. 1966. *The Meanings of the Modals in Present-Day American English.* The Hague: Mouton.

Hopper, P. J. 1991. "On some principles of grammaticization". In *Approaches to Grammaticalization, Volume 1: Focus on Theoretical and Methodological Issues.* E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam and Philadelphia: John Benjamins.

Hopper, P. J. and Traugott, E. C. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Jespersen, O. H. 1961. *A Modern English Grammar on Historical Principles: Part II: Syntax.* VII. London: George Allen & Unwin.

Krug, M. 1998. "*Gotta* — the tenth central modal in English? Social, stylistic and regional variation in the British National Corpus as evidence of ongoing grammaticalization". In *The Major Varieties of English*, H. Lindquist, S. Klintborg, M. Levin and M. Estling (eds.), 177–191. Växjö: Växjö University Press.

Labov, W. 1972. *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Labov, W. 1982. "Building on empirical foundations". In *Perspectives on Historical Linguistics*, W. P. Lehmann and Y. Malkiel (eds.), 79–92. Amsterdam and Philadelphia: John Benjamins.

Mitchell, B. 1985. Old English Syntax. 1–11. Oxford: Clarendon Press.

Palmer, F. R. 1979. *Modality and the English Modals.* New York: Longman.
Poplack, S. and Tagliamonte, S. 2001. *African American English in the Diaspora: Tense and Aspect.* Malden: Blackwell Publishers.
Rand, D. and Sankoff, D. 1990. *GoldVarb: A Variable Rule Application for the Macintosh.* Montreal, Canada: Centre de recherches mathématiques, Université de Montréal. Version 2.
Tagliamonte, S. A. 1998. "*Was/were* variation across the generations: View from the city of York". *Language Variation and Change* 10 (2): 153–91.
Tagliamonte, S. A. 2002. "Comparative sociolinguistics". In *Handbook of Language Variation and Change*, J. Chambers, P. Trudgill and N. Schilling-Estes (eds.), 729–763. Malden/Oxford: Blackwell Publishers.
Tagliamonte, S. A. 2003. '"Every place has a different toll': Determinants of grammatical variation in cross-variety perspective". In *Determinants of Grammatical Variation in English.* G. Rhodenberg and B. Mondorf (eds.), Berlin and New York: Mouton de Gruyter.
Traugott, E. C. 1997. "Subjectification and the development of epistemic meaning. The case of *promise* and *threaten*". In *Modality in Germanic Languages. Historical and Comparative Perspectives*, T. Swan and O. Westvik (eds.), 185–210. Berlin: Mouton de Gruyter.
Trudgill, P. 1990. *The Dialects of England.* Oxford: Blackwell Publishers.
Visser, F. T. 1963–73. *An Historical Syntax of the English Language.* Leiden: E. J. Brill.
Warner, A. 1982. *Complementation in Middle English and the Methodology of Historical Syntax.* London and Canberra: Croom Helm.
Warner, A. 1993. *English Auxiliaries: Structure and History.* Cambridge: Cambridge University Press.

The semantic path from modality to aspect

Be able to in a cross-linguistic perspective*

Karin Aijmer
Göteborg University

Ability has a number of lexical sources including dynamic, telic verbs of various sorts. For example, *manage to,* a verb implying that there were difficulties to overcome can come to mean 'ability', a development which has parallels in other languages. In this paper translations into Swedish are used as a window to show that *can*, *be able to* and *manage to* are closely related to each other. It is argued that in grammaticalization studies of modality one has paid a lot of attention to 'ability' as a source for deontic and epistemic meanings but neglected its relevance as a source for the development to an aktionsart meaning associated with actuality and result.

> In its broadest interpretation, the goal of linguistics is to discover how human languages are alike and how they differ, and to propose and test theories that explain the similarities and differences. There are many alternative paths to this goal, for there is much to explain. (Bybee et al. 1994: 1)

1. Introduction

Be able to has often been treated in a stepmotherly way as a synonym of *can* which must be used because the *can/could* paradigm is defective. But it also needs to be analysed in its own right. *Be able to* can, for instance, be regarded as a lexical source from which new meanings can be derived in a principled way rather than as a substitute for *can/could.*

This article was undertaken with the aim to investigate verbs and related constructions with the meaning dynamic possibility or ability in English and Swedish and their semantic developments into the realm of actuality. The paper could also be called 'how to be able and how to succeed in English and Swedish'

since both ability and success will be dealt with as sources of modal and aspectual meanings. In the literature on grammaticalization much attention has been given to the semantic development from ability to deontic and epistemic possibility (cf. e.g. Traugott 1989, Van der Auwera and Plungian 1998). When one studies modality and its realisations one also finds a development from ability to actuality.

The semantic developments which can be postulated on the basis of only two languages could be regarded as a hypothesis about a general development which can be confirmed or rejected on the basis of data from more languages. Traditionally linguists have asked native informants to make judgements about meanings. Native speakers can distinguish different uses of the same polysemous item and say how they are related. However, at times too many uses may be distinguished; at other times too few (cf. Bybee et al. 1994: 44). Translations are more accessible and reliable as sources of meaning and uses than native informants since they are produced by trained translators without any theoretical concerns in mind (Noël, forthcoming).

Following Dyvik (1999) we can regard the translational properties of a linguistic sign as a window onto its meaning properties and not only as the result of a matching of words from different languages. In this article I describe the meaning and uses of the English verbs *be able to* and *manage to* on the basis of their occurrences and translations in the English-Swedish Parallel Corpus. The Swedish verb *lyckas* is a frequent translation of both *manage to* and *be able to* and will also be compared with English translations. The English-Swedish Parallel Corpus consists of almost three million words of fiction and non-fiction representing a number of different genres. The texts are English originals and their translations into Swedish and a corresponding number of Swedish texts representing the same genres and their translations into English (for more information on the corpus, see Aijmer et al. 1996, Altenberg and Aijmer 2000). The methodology of using translation corpora for cross-linguistic studies is still new and complements methodologies in language-typological studies where data-gathering is based on published reference material or questionnaires.

A demonstration that a form has certain meanings or uses on the basis of translations is not an explanation of why we find these correlations. Following recent suggestions in cross-linguistic typological studies that the different meanings expressed by modality in a language can be represented by a number of modal paths I will discuss the developments from the lexical source potentiality/ability and success.

2. Ability and actuality — theoretical prolegomena

In traditional descriptions of modality, subject-oriented 'ability' is said to be concerned with dynamic ability (von Wright 1951: 8) but it is often omitted from typological classifications of modality since it is not concerned with the opinion or attitude of the speaker (Palmer 1986: 103). In this paper ability will be regarded as a type of possibility (referred to below as 'participant-internal possibility') from which new meanings can be created by extension. Ability can be described by the following factors (cf. Leech and Coates 1980: 83f). The factors can be exemplified with *can* and be referred to by semantic features:

'the ability *can*' ('participant-internal possibility')
[+agentive] — the subject-referent has agentive function
[+possibility] — there is nothing which prevents the subject from doing X
[+inherent properties] — the possibility of the action is determined by inherent properties (whether physical or psychological capabilities) of the subject-referent
[+activity] — the main verb denotes an action or activity

In the example *John can swim*, there is nothing that prohibits the subject-referent from doing X. The possibility of the action taking place is determined by the inherent abilities (or physical capacity) of the subject-referent.[1] When the enabling factors are the subject's inherent capacity *can* can be referred to as participant-internal modality (possibility) (Van der Auwera and Plungian 1998: 80). If the subject is inanimate or the circumstances are less ability-like *can* is less securely assignable to the ability category (cf. Van der Auwera and Plungian 1998: 80 'participant-external possibility').[2] Coates illustrates this with the example:

> Britain's word can still be of value in some parts of the world.
> (Coates 1983: 90)

In the case of possibility we can refer to circumstances of diverse kinds as enabling factors. Ability, on the other hand, refers to participant-internal possibility. This has as result a difference between *can/could* and *be able to*. Cf. Klinge (1993: 333) who claims that "*be able to* assigns an inherent quality of power to the referent of the subject noun phrase, which cannot be subject to the wishes of the addressee" (quoted from Facchinetti 2000: 125). For example in (1), the modality is asserted (the inherent ability of the subject to do something):

(1) He is able to create enthusiasm, which helps the electorate no end.
(example slightly adapted from Facchinetti 2000)

If I say that somebody is able to create enthusiasm I am predicating the subject's ability to do this because of factors inherent in the subject himself. However it is possible to assert (1) but deny it afterwards: "He is able to create enthusiasm but in his present job he has no chance to do it". In other words, the implication has not been conventionalized. In "he can create enthusiasm", the ability is generic and there is no implication of actuality. *Be able to* but not *can* can be used to predict something about a future event. In the example below, *can* is blocked and *be able to* must be used:

I'll be able to swim by this time next year.
(Edmondson et al. 1977:311)

Consider now the propositions expressed by the assertions in (a) and (b):

(a) She was able to finish her dissertation.
(b) She finished her dissertation.

In an interesting paper, Panther and Thornburg (1999:334) suggest that there is a "metonymic link" between (a) and (b), i.e. statement (a) can be used to "stand for" statement (b), the difference being that in (a) the speaker predicates the ability to finish the dissertation (and implies that she will finish it), whereas in (b), the speaker predicates the actuality of finishing it. In pragmatic terms, (b) is a (generalized) pragmatic implicature induced by (a) (Panther and Thornburg 1999:335). Compare also Facchinetti (2000:118): "within the value of ability, *be able to* is generally acknowledged to have an implication of actuality, which *can* lacks."

Edmondson et al. in their pedagogical grammar of the English verb regard *manage* as a verb which may replace *be able to* in the ability meaning (1977:311). Their example is:

On one occasion I managed to beat Larsen at chess.

The implication is that the speaker actually beat Larsen at chess and it cannot be cancelled. The sentence would be ungrammatical with *could*:

*On one occasion I could beat Larsen at chess.

What this shows is that *could* does not imply actuality although *be able to* does (Palmer 1977:5). On the other hand the speaker could say *I can beat Larsen at chess* which would imply that the event may take place at some time in the future.

Cf. Palmer (1977:5), "If [...] a modal cannot be used in a non-modal context (i.e. where the 'factual' status is known), this will exclude the use of *could* to suggest the actuality of a single event in the past, but not the use of *can* to suggest the actuality of an event in the future."

According to Facchinetti, "the idiom [*be able to*] appears to be gaining ground when the emphasis is on the inherent quality or power of the subject of the utterance, who 'manages' or 'does not manage' to carry out an action" (2000:128). For example, instead of (a) or (b) the speaker could have used (c) which shows the actuality meaning of *be able to*:

(c) She managed to finish her dissertation.

There is however no logical reason why *be able to* should have this property but not *can/could*. In Swedish we can for instance use *kunde* 'could' as a correspondence of (a) or (c) indicating that *kan/kunde* and *can/could* has a different status in English and in Swedish although they are cognates (cf. Aijmer 1999). Consider (2) where a form of *kan* corresponds to *manage*:

(2) Journalisterna hade aldrig *kunnat* besegra honom. (HM1)
The reporters had never *managed* to beat him.

The similarities and differences in meaning between *can, be able to* and *manage* are summed up in the following feature matrix:

	Agentive	**Activity**	**Possibility**	**Inherent enabling factors**	**Effort**[3]	**Actuality**
can	+	+	+	+/−	−	−
be able to	+	+	+	+	−	+/−
manage	+	+	+	+	+	+

It is claimed by Panther and Thornburg (1999) that the "Potentiality to Actuality" metonymy is exploited in different languages and in different conceptual domains (such as sense perceptions and actions). However, in saying "she was able to finish the dissertation" or "she managed to finish the dissertation" the speaker is not saying the same thing although the implication would be the same. "She was able to finish the dissertation" would be used if the speaker explicitly wanted to focus on the effort or successful attempt to carry out the action if it has been suggested that there were some difficulties.

To sum up, the speaker uses *be able to* with action verbs to focus on the

subject's ability to complete an action, the attempt and effort to achieve a successful result. *Be able to* implicates actuality (the successful result of the ability) just like *manage to*. It follows that "John was able to read the book" implies that John actually read the book just as "He had learnt how to swim" implies that he knew how to swim and that from "they like to go for walks" we can infer that they actually go for walks. In pragmatic terms we have conversational implicatures which have been generalized, since they do not require particular contextual conditions or a special scenario in order to be inferred (Levinson 1983:126).[4]

Edmondson et al. (1977) have suggested that the reference to a specific occasion or event is important for when we use *can* or *be able to*. The evidence from the translations suggests that another factor is the aktionsart of the verb which is governed by *be able to* or by *manage*.

In what follows I want to further explore when the speaker uses *can*, *be able to* or *manage to* and what the implications are of the different choices. In this paper translations into Swedish will be used as a window to find out how *can, be able to* and *manage* are used and what they mean.

3. *Can, be able to* and *manage to* in English-Swedish contrast

The translation method reveals alternative ways of translating the same lexical item e.g. as a lexical verb or as a modal auxiliary. The translations can be arranged as a paradigm where the items correspond to different meanings of the source item.

I have little to say about *can/could* in general since I have only looked at the cases where it was translated by *lyckas*. In the negative the difference between *can* and *be able to* is neutralized for example in "he couldn't get back" which implies that he didn't (manage to) get back (Palmer 1977:6). Another context is illustrated in (3) where success is only hypothetical:

(3) If we *can* get a scab off a finger we put the finger under and watch the place where the blood oozes out, bright red, in a round button, like a berry. (MA1)
Om vi *lyckas* få loss en sårskorpa från ett finger sätter vi fingret inunder och betraktar stället där blodet sipprar fram, klarrött, i en rund knapp, som ett bär.

Both *be able to* and *manage to* are interesting from a translation perspective. In total there were 70 examples of *be able to* in the English originals and almost

twice as many examples in the English translations of the Swedish sources. The Swedish translations are shown in Table 1a (Appendix) and include both a modal verb (*kan*, *kunde*) and lexical verbs. *Kan*, *kunde*, *kunna* was used as a translation in 71% of the examples. The cases where the translator has not chosen a modal verb are of particular interest as evidence of the function of *be able to* and *manage to*. The ability reading is also indicated in the translations *bli i stånd*, *kapabel att* 'become capable of'. The translation by *lyckas* ('manage to') was found in 11.4% and a simple or complex verb (with omission of *be able to*) in slightly fewer cases (10%).

The Swedish sources of *be able to* are given in Table 1b. The sources may be specific verbs such as *orka* ('have the physical or psychological strength to do sth'), *hinna* ('have the time to do sth') making the enabling conditions explicit. Omission (zero-correspondence) is of interest since it provides evidence that actuality is implicated by the verb. In (4), a specific action is implied by the use of *be able to*:

(4) If you put me down at the crossroad he'll be able to give me a lift. (PDJ1)
Om ni släpper av mig i korsningen så kör han mig hem.

When *kunde* was used in the translation, ability was asserted. It is not always easy to distinguish between ability and success in actual discourse. In (5), *kunde* 'could' is not replaceable by 'lyckades' suggesting that the focus is on the ability rather than on the attempt and effort to achieve a result:

(5) 'What exactly would you say this might be?'
'Tomato,' I *was able to* inform her. 'Skinned, cored, depipped, cubed. (JB1)
"Vad, närmare bestämt, skulle Oliver säja att det där är?"
"Tomater", *kunde* jag upplysa henne. "Skållade, skalade, kärnade — och tärnade."

However in (6), *be able to* was translated by *lyckas* signalling that actuality is more strongly implied:

(6) I *was able to get* some Lotromycin. (AH1)
Jag *lyckades skaffa fram* litet Lotromycin.

The difference between (5) and (6) seems to have to do with the degree of effort involved in carrying out the action expressed by the verb rather than with actuality. Whatever the "before" conditions are they could be regarded as subcases of the ability condition (Panther and Thornburg 1999: 338). For a simple verbal action like 'opening the window', the conditions that the subject

has the strength, the patience, the time to open the window would normally be taken for granted. Similarly, giving information (having the required expertise) has other felicity conditions than getting something from the pharmacy which may imply first getting a prescription, going to the pharmacy, etc.

These examples show that the type of event is important. Events are said to denote accomplishments (rather than activities) when the process expressed by the verb has a logical culmination. The following classification which has been suggested by Pustejovsky implies that verbs and verb phrases differ in the kind of eventualities in the world they denote (Pustejovsky 1995: 12). As a result we can distinguish three aspectual types or classes ('aktionsarten'): state, activity and event. States and activities are of less interest when the verbs have an inherent actuality reading.

Below I look particularly at examples where *be able to* refers to accomplishments since this is a context where the verb has developed new meanings. In (7), *be able to* is translated by *lyckas* implying that there were difficulties to overcome:

(7) Vic and Roger had wanted it to read I DON'T SCREW AROUND, but they hadn't *been able to* sell the Voit people on that. (SK1)
Vic och Roger hade velat skriva INGA SKITGREJER, men de hade inte *lyckats* få Voitfolket med på noterna.

The correspondences of *manage to* in Swedish could be more specific since they mention the enabling factors (Table 2a). *Hinna* ('have time to') was especially frequent in sources (Table 2b):

(8) Jag *hann* lära mig en del… *Hon* lärde mig en del. (GT1)
I *managed to* learn a bit myself… *She* taught me a bit.

4. Zero-correspondences

What kinds of verbs and constructions are involved in the governed clause when we get zero-correspondences? I would like to pick out for discussion some verbs and constructions as equivalent to *be able to* or *manage to* which do not come immediately to mind but turn up in translations and show the complexity of describing the meaning of ability verbs. The same meaning as *be able to* (*manage to*) and its complement can be expressed morphologically (in some languages), by particle verbs (*put on*), by verbs (*get* in different functions as well as related verbs such as *catch, persuade*), causative verb anticipating the logical culmination of an event (cf. Brinton 1988).

Get is a typical example of "participant-external actuality", "meaning that some state of affairs is actualized because of participant-external circumstances, in particular when the participant in some way 'received' the state of affairs" (cf. van der Auwera & Plungian 1998: 105). However *get* expresses "participant-internal actuality" in certain syntactic frames. Examples such as (9) imply that the subject made an active attempt and succeeded in doing something because there is logical culmination to the event:

(9) From there *he was able to get* a different view of the woman opposite (RR1)
Därifrån *fick* han en annan infallsvinkel till kvinnan mittemot

When *manage* in the English translations corresponds to *få* (or an idiom with *få*) in the Swedish source text, *få* (like *get*) has the meaning 'internal actuality' (cf. van der Auwera & Plungian 1998):

(10) Jag snubblade till, men *fick fatt i* en lina, gjorde fast och gick sedan akteröver och stängde av motorn. (BL1)
I stumbled, but *managed to find* a line, make fast and go astern to turn off the engine.

The Swedish original in (10) is translated by "managed to find a line" rather than simply "found a line" (cf. "I stumbled, but got hold of a line"), i.e. *manage to* is needed when there is a focus on ability and attempt to do something. In Swedish the same meaning is expressed by the verb+particle construction.

In (11), the Swedish translation has a particle verb (*få loss*) corresponding to *be able to* in the English source:

(11) Hon *fick loss* lädret till slut men då såg käften likadan ut som när hon bitit till. (KE2)
Eventually she *was able to extract* the leather, but her jaws remained clamped.

In Swedish, *få* in the participant-internal reading has low frequency (cf. Viberg 2002: 137). More frequently *få* (as a translation of *be able to*) has modal (deontic) meaning:

(12) Det är ju väldigt värdefullt att man *får* koncentrera sig på en del och göra färdig den. (CE1)
It's tremendously valuable, *being able to* concentrate on one thing and finish it.

Particle verbs have aktionsart meaning and express accomplishment: *hold on to*, *put down*, *stave off*, *prise off*, *burst out of something*, *wear sb down*. In (13) a particle verb in English is translated with *lyckas*:

(13) Joseph *got up* and, empty-handed, ran up the flight of stairs, into the light and safety of the street.(JC1)
Joseph *lyckades resa sig* och sprang tomhänt uppför trappan och ut mot ljuset och tryggheten på gatan.

Other verbs with aktionsart properties are causative verbs. When the translation has a causative verb or a causative construction, omission of *manage to* could be explained by the wish to avoid a doubling of function. *John managed to persuade him to do it* sounds clumsy (in English) in comparison with *John persuaded him to do it* since both *manage* and *persuade* imply the termination of the activity. Similarly informants would say that *managed to* is not necessary since the same meaning is expressed by the type of verb (idiom, verb plus particle).

No correspondence in sources was more frequent than zero-translations. *Able to* has been added in the English translations from the Swedish sources in 14% of the cases metonymically implying that something is the case. The corresponding figure for *manage to* is 20% (see Table 2b). *Be able to* has often been added when the Swedish original has the present perfect or past perfect. In Swedish the perfective form expresses the effect of a successful action. *Be able to* has been added in examples (14) and (15):

(14) Om hon *hade sett* honom i dag hade hon haft annat att klaga över. (LG1)
If she *had been able to see* him today she would have had other things to complain of.

(15) Egentligen *har jag aldrig erbjudit* någon en känsla på riktigt. (MS1)
As a matter of fact, I'*ve never been able to offer* anybody a genuine feeling.

5. Swedish *lyckas* and its correspondences in English

Swedish *lyckas* corresponds both to English *succeed* and to *manage to* (to be successful in the endeavour to do something). It is both intransitive ('be successful') and a catenative verb followed by an infinitive (corresponding to 'manage to do something'). The semantic development into the domain of actuality we have described is not only an interesting linguistic fact about

English. It occurs in Swedish as well, suggesting that this is a general tendency explained by what goes on in conversation.

The paradigm in Table 3a shows the correspondences of Swedish *lyckas* in the English translations (73 examples in all). The equivalence between Swedish *lyckas* and *manage* or *succeed* is expected since the lexical (semantic) source is the same. The relative frequency (54.8% of the total) shows that *lyckas* is often translated by a word with the same meaning in English. Neither *able to* nor *could* were frequent. In 16.4% of the examples there was no correspondence in the English translations (zero-correspondence).

The possibility of using a verb 'metonymically' to refer to something actual was exploited in Swedish with the lexical verbs *lyckas, orka, hinna* (to give a few examples).

When *lyckas* is translated by *manage* there may be something in the context indicating that there is focus on the attempt and the effort to do something:

(16) Efter visst besvär *lyckades* de få tag på en bandspelare som passade. (HM2)
With some considerable difficulty they eventually *managed* to find a machine to play it.

The linguistic cues to the success interpretation are "with some considerable difficulty" and "eventually".

Could occurs as a translation of *lyckas* in certain contexts only. When the context is negated, *could* implies actuality (17) and can be a translation of *lyckas*:

(17) Han *lyckades* inte motstå frestelsen att trycka till lite extra med en av knivarna mot fingertoppen. (HM1)
He *couldn't* resist the temptation to press a little harder on one of the knives with his fingertip.

Could also occurs in an *if*-clause with the implication of hypothetical success.

(18) Men om man *lyckades* lirka ur Gunnar ett tillstånd skulle det sedan gälla. (JMY1)
But if you *could* talk Gunnar into granting permission, it would stick.

When *lyckas* is translated by *be able to*, this indicates that *be able to* has developed the meaning ability:

(19) Möjligen kunde fortfarande någon på den tiden ung medlem i Skogskrematoriets byggnadskommitté vara i livet, men inte ens den alltid lika energiskt intresserade och vänligt hjälpsamme Börje Olsson på Stockholms kyrkogårdsförvaltning *lyckades* hitta någon. (CE1)

Perhaps somebody who at that time had been a young member on the Woodland Cemetery Building Committee might still be alive, but not even the indefatigably enthusiastic and helpful Börje Olsson of the Stockholm Cemeteries Authority *was able to* trace anyone.

(20) Eftersom man inte *lyckades* bemästra de sanitära problemen avlöste den ena epidemien den andra. (KF1)
Since nobody *had been able to* solve basic sanitary problems, one epidemic followed another.

When the translator has not translated *lyckas* the reason is often that it would have been clumsy to use *manage* (*could, was able*). In (21), the translator has chosen transposition to the passive in order to avoid *manage* in the translation:

(21) Det brann ner fullständigt och *alla lyckades rädda sig* utom en barnpiga och en liten femårig flicka som låg i hennes rum. (MR1)
It burnt down to the ground and *everyone was saved* except a nursemaid and a little five-year-old girl who had slept in her room.

When *lyckas* is negated it can easily be omitted:

(22) Han ryckte till av att Hansson sa något som han *inte lyckades* uppfatta. (HM1)
He gave a start. Hansson had said something he *didn't catch.*

In (23), it would hardly have been acceptable to say "he managed to manufacture" a saddle:

(23) Förvånansvärt snabbt *lyckades* han tillverka en sadel till varje åsna.(LH1)
In a surprisingly short time, he *manufactured* a saddle for each and every one.

In (24)–(26) the effort and attempt are sufficiently expressed by the particle verb (*hold on*, *put down*) or by the causative verb (*persuade*) (cf. Brinton 1988).

(24) På Island *lyckades* ett mindre antal ledande släkter hålla på de flesta — om än inte alla — viktiga ämbeten. (HG1)
In Iceland a small number of leading families *held on* to most — but not all — important posts.

Managed to hold on is not necessary in the translation since the meaning of capacity, effort and attempt is expressed by the particle verb.

In (25), *put down* is sufficient to express the accomplishment meaning expressed by *lyckas* in the original:

(25) Enligt gamla men historiskt osäkra kungalängder skall det ha varit den nionde kungen efter Olof Skötkonung, Inge den äldre, som till sist *lyckades* slå ner det sista asaupproret, lett av hans egen svåger, Blot-Sven kallad. (HL1)
According to ancient and historically uncertain royal records it was the ninth king after Olof Skötkonung, Inge the Elder, who *put down* the last Aesir revolt which was led by Inge's own brother-in-law, known as Blot-Sven.

Persuade as a causative verb has accomplishment meaning:

(26) Efter en stund hade Stanley *lyckats* förmå honom att långsamt, steg för steg, redogöra för sina ansträngningar att få fram bärare. (LH1)
After a while, Stanley *persuaded* him to give a step-by-step report on his efforts to procure porters.

Get, as expected, was used as a correspondence to *lyckas/lyckades* + verb:

Swedish	**English**
lyckas närma sig 'manage to approach'	*get close*
lyckas få av höljet 'manage to get the wrappings off'	*get the wrappings off*
lyckas resa sig 'manage to rise'	*get up*
lyckas utverka ett löfte 'manage to get a promise'	*get a promise*

When we look at the sources of *lyckas* in Table 3b, there was no direct correspondence in 35.5% of the cases. The English source texts have simple and complex verbs expressing accomplishment (idioms, causative verbs, particle verbs, the verbs *get* and *make*, "creation verbs"). The translator has used a verb such as *manage* or *be able to* even when it would have been possible to find a corresponding idiom, particle verb, etc. Such differences between translations and sources are not unusual when the language systems which are compared are different. They show that it is important to look at both translations and sources in a translation corpus. It is also well-known that there are differences between languages in the areas of modality and aspect. These differences are however subtle but clear when two language systems are confronted with each other.

6. *Be able to* and *manage* as sources of grammaticalization

Constructions such as *be able to*, *have to*, etc., have long puzzled grammarians because they resemble modal auxiliaries without fulfilling all the criteria for

auxiliaries. Biber et al. (1999: 484) treat *be able to* (and, for example, *want to*, *be obliged to*, *be willing to*) as relatively fixed expressions or "lexical bundles" with meanings similar to the modal auxiliaries (Biber et al. 1999: 484, 1022).

It may be thought that such "semi-modals" came into existence when "the reanalysis of pre-modals created a vacuum which was immediately filled by a new set of semantically equivalent verbs" (cf. Brinton 1988). A similar view on *be able to* is found in other work as well. For example, Quirk et al. (1985: 143) refer to *be able to* and related constructions as semi-auxiliaries and point out that they can fill slots in a modal verb paradigm (p. 144).

However, neither diachronically nor synchronically is there any need to see a new or alternative marker as contingent on the loss or dysfunction of another marker (cf. Bybee et al. 1994: 21). For example, there is no reason to believe that native speakers experience *be able to* as closely related to *can* and *could*. Brinton makes the same point with special reference to "quasi-modals":

> The "quasi-modals" represent a second stage of grammaticalization of full verbs — verbs already existing in the language, not newly created — as auxiliaries of mood. They are selected for grammaticalization because of their semantic equivalency to modals, but because of the later onset of grammaticalization, they have not yet acquired all the syntactic features of auxiliaries which follow from their semantic function. (Brinton 1988: 106)

As pointed out by Brinton, verbs like *be able to* and *manage* deserve attention from the point of view of grammaticalization. Following the pioneering work by Meillet, grammaticalization can be defined as the development of lexemes into grammatical items (Traugott & Dasher 2002: 81). Subsequent work has suggested that grammaticalization should be conceived more specifically as "the change whereby lexical material in highly constrained pragmatic and morphosyntactic contexts is assigned functional category status" (Traugott & Dasher, ibid.).

There has been much discussion about the characteristics of grammaticalization, such as the reanalysis from open class to a closed class of items and the fixing of constructions. *Manage to* and *be able to* have few syntactic features associated with auxiliarihood, i.e. the meaning is not coded in a special linguistic form. However the role of syntax and 'structuralization' for grammaticalization is not uncontroversial. Nuyts (2001: 177) claims that "although it is true that meanings expressed by grammatical markers are usually of the qualificational type [tense-aspect-modality type] such meanings are also frequently expressed by purely lexical categories". The development of modal and aspectual meaning could be seen in itself as a sign of grammaticalization. In an

approach to grammaticalization focussing more on semantic developments, syntactic changes are language-specific or even accidental. Brinton similarly claims, for example, that "the syntactic changes which follow a word's reanalysis as an auxiliary are 'accidental' in that they depend upon the age, frequency of occurrence, and importance of the markers" (1988: 110).[5]

The literature on grammaticalization has mostly dealt with the mechanisms of change (such as metaphor and metonymy) or the typology of widely attested semantic-pragmatic changes, as in Hopper & Traugott's work (1993). The shifts undergone by *be able to* and by *manage* are, for example, changes in the lexicon not unrelated to the famous change from 'bead' (prayer) to the balls of a rosary, a change which Stern (1968) called permutation but which is identical to the type of metonymy discussed in more recent work focussing on the mechanism of conceptual and discursive metonymization (Traugott & Dasher 2002: 67).

Lately grammaticalization studies have taken a different direction. In cross-linguistic typological studies (Bybee & Pagliuca 1985, Bybee et al. 1994, van der Auwera & Plungian 1998) attention has been drawn to lexical source as a basis for building semantic accounts of grammaticalization. For example, van der Auwera & Plungian (1998) propose that grammmaticalized expressions of modality can be supplied with a semantic map in which synchronic and diachronic connections between modal, premodal and post-modal meanings and uses can be shown. In a map nothing is implied about order or temporal priorities and no language-specific developments are referred to.

The notion of grammaticalization path (Traugott & Dasher 2002: 86) is a similar way of describing a diachronic pattern representing regularities which can be observed over time. The categories in the path do not shift, but the path represents the probable new polysemies of a form if it changes meaning. Observed prototypical shifts are replicated across languages and can also be represented as paths (or as semantic maps). Thus in many European languages including English and Swedish the meanings of possibility modals are typically both deontic and epistemic (cf. Aijmer 1999).

Epistemic possibility ('it is possible that') refers to a judgement of the speaker about the certainty or uncertainty of the proposition and can be related to ability, permission and 'root' possibility. Neither permission nor epistemic possibility will be further discussed here since they cannot be expressed by *be able to*. In this article I have focused instead on a semantic path which has been less often discussed. *Manage to* (success) and *be able to* (ability) are both good examples of sources from which modal and aspectual meanings can be derived (cf. Bybee et al. 1994). The change undergone by *manage to* could, for example,

be described in terms of a map representing the development from successful completion to ability with a loss of the feature 'effort' (Bybee et al. 1994: 190).

The framework in which semantic changes take place involves pragmatics (rather than mechanisms such as metaphor). Hearers construe inferences "on the fly" which may then be exploited for different purposes or strategies. Such inferences can come to be conventionalized and be present as polysemies of the verbs. *Can* has, for instance, the meanings ability, possibility and permission. Ability can also come to imply actuality as has been demonstrated for *be able to*. The metonymic link between potentiality and actuality explains that if the subject is able to do something, the hearer is allowed to make the inference that this will have the realization of an event as the result.

There are many possible semantic paths which are only constrained by the lexical source. Participant-internal potentiality or ability plays an important role as a source for semantic developments. If we say "he was able to open the window" or "he managed to open the window" it is implied that there was some obstacle which had to be overcome. This implication needs to be explained with reference to metonymy (potentiality can stand for what is actual) and/or by conversational implicature (Traugott & Dasher 2002 "invited inferences") and conversational maxims. There must be a special reason for saying explicitly that 'he was able to open the window' which the hearer can be supposed to infer from the context. Our knowledge of what is involved in opening a window provides a scenario which is exploited for drawing inferences. When the speaker states that he was able to do something more than actuality is implied, namely that there was an obstacle to overcome and the action is the successful result of an effort to overcome the obstacle.

Manage to has not developed the meaning possibility more than in a few contexts but has developed aktionsart meaning while *be able to* is either modal or an aktionsart marker. Success or achievement is also conveyed by the aktionsart of the verb. It follows that *be able to* or *manage to* can come to be redundant. The large number of zero-translations where the governed verb or the context takes over the responsibility for expressing result demonstrate nicely the importance of syntagmatic context for semantic change.

7. Conclusion

Ability has a number of lexical sources including dynamic, telic verbs of various sorts. For example *manage to*, a verb implying that there were difficulties to

overcome can come to mean 'ability', a development which has parallels in other languages (cf. Bybee 1994: 191). Since 'ability' can develop into 'possibility', *manage to* may also receive this meaning. When *manage to* corresponded to *kan, kunde* in Swedish this indicates that the verb has got modal meaning and that there is a close relation between success, ability and possibility. *Be able to* was sometimes translated by *få* ('get') suggesting that 'get', 'obtain' is another source for ability (cf. the development *get* > 'able', *get* > 'manage' in some languages parallel to the Swedish development (Heine & Kuteva 2002: 143; Bybee et al. 1994: 190f).

A further development involves the grammaticalization of ability to an aspectual marker. This should not be explained as semantic weakening but as a metonymic development from participant-internal ability to participant-internal actuality in certain contexts. Evidence for this development comes from zero-translations. *Manage to* has been grammaticalized directly from a lexical element to a grammatical marker. The development from ability to successful completion may be an instance of a more general pathway whereby process verbs give rise to markers of tense, aspect, and modality in certain contexts. Heine & Kuteva (2002: 148f) mention, for example, *begin*; *come from*; *come to*; *do*; *finish*; *get* > *obligation*; *get* > *permissive*; *go to*; *keep*; *leave*; *put*. The aktionsart meaning of 'ability' is derived in collocations with certain verbs (accomplishment verbs) and involves semantic bleaching.

To what extent have the lexical verbs also been become auxiliaries? Modal auxiliaries have to varying degrees a deficient verbal paradigm and they have lost their ability to be inflected for features like person and number. Formally *manage to* (*lyckas*) and *be able to* are more like lexical verbs than auxiliaries. The verbs are lexical because they are inflected for tense and person. They may be preceded by another modal auxiliary (*can manage*, *must be able to*). Functionally both *manage* and *lyckas* are similar to auxiliaries since they have grammatical rather than lexical meaning (inherent ability). *Lyckas* is more grammaticalized than *manage* in English as indicated by the fact that it was almost regularly inserted as an 'aspectualizer' before certain verbs in translations into English. When used as an auxiliary, it indicates that the action has taken place. Its auxiliary status is also indicated by *to*-deletion (Swedish *att*).

Can, could on the other hand has developed the participant-internal actual meaning in certain contexts (negative, hypothetical) where it was translated by *lyckas*. Unlike *be able to* it has however acquired deontic meaning and (in some languages) the meaning of epistemic possibility.

The hypothesis about how ability, possibility and success are related can be

tested by bringing in more languages. English and Swedish are genetically related languages and similar tendencies can be expected. It is interesting to note that the Romance languages also use success verbs as auxiliaries with the meaning 'ability' or 'possibility'. French *réussir* and Italian *riuscire* are found in contexts where Swedish uses *kunna* or another modal verb to signal ability (Lindvall 2002). There is a short step to an 'aspectualizer' which is optional when the governed verb indicates result. *Riuscire a* (participant-internal actuality) in Italian would, for instance, be used whenever a result can be referred to an inherent property. However we need to be careful with cross-linguistic generalisations in the area of modality. As Palmer points out (1977: 21):

> Ultimately, much of what happens in a language is a matter of the brute facts of the language. The explanations that have been suggested are very largely not explanations that say that things could not have been otherwise. Knowing all the relevant facts we could not have predicted what would actually occur — we could not have predicted, for instance, the fact that *could* does not imply actuality while *was able to* (and *potui* and *ai pu*) does. This, of course, raises the whole problem of what is 'possible' in language and so what in language is specific and what is universal because it must be so.

Languages do not necessarily make the same semantic distinctions, as is clear from this little investigation. However, the existence of similar distinctions or developments in many languages shows that they have some rational basis in what happens in the communication situation. In this light we can explain that there is a similar tendency in languages for changes from ability to possibility, permission and actuality. Ability can itself have several sources (e.g. 'manage' or 'get') which may partake in the developments undergone by ability.

Notes

* Many thanks to Bengt Altenberg and to Göran Kjellmer for reading an earlier version of the manuscript. The paper has also profited from fruitful discussions with Lars Lindvall who made me interested in the Swedish verb *lyckas* and its correspondences in other languages.

1. Including learnt ability.

2. Cf. also Bybee et al. (1994: 290) on the change from (mental) ability to general ability through the loss of features of meaning.

3. The feature 'effort' would also be used to distinguish between *try* and *manage*.

4. Cf. Traugott & Dasher (2002: 35) "generalized invited inference".

5. Traugott & Dasher (2002: 87), on the other hand, regard the semantic changes observable in grammaticalization as a subtype of regular changes in the lexicon:
"What is different is that in grammaticalization semantic changes affect only certain Ls [lexemes] (those with relatively general meaning) and correlate with other changes such as morphosyntactic decategorialization."

Appendix

No distinction has been made in the tables between different forms of the verbs.

Table 1a. Translations of *be able to* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
KAN[a]	50	71
lyckas ('manage to')	8	11.4
Ø	7	10
blev i stånd ('become capable')	1	1.4
förmå ('enable')	1	1.4
kapabel	1	1.4
Other	2	2.9
Total	70	

[a] The lemma KAN includes *kan, kunde, kunnat, kunna*

Table 1b. Sources of *be able to* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
KAN	102	71
Ø	20	14.1
lyckas	6	4.2
orka ('have the strength to')	4	2.8
få ('get')	4	2.8
hinna ('have the time to')	3	2.8
transposition Active → Passive	2	2.1
gå att ('go to', 'it is possible that')	1	0.7
Total	142	

Table 2a. Translations of *manage to* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
lyckas	31	75.6
Ø	7	17.1
råka ('happen to')	1	2.4
KAN	1	2.4
orkar ('have the strenght to')	1	2.4
Total	41	

Table 2b. Sources of *manage to* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
lyckas	40	47.6
Ø	17	20.2
KAN	6	7.1
orka	7	8.3
hinna	14	16.7
rå med att ('have the capability to')	1	1.2
Total	84	

Table 3a. Translations of *lyckas* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
manage	40	54.8
succeed	10	13.6
Ø	12	16.4
could	2	2.7
able	2	2.7
transposition active → passive	2	2.7
successfully	1	1.4
could manage	1	1.4
be successful	1	1.4
unable	1	1.4
be lucky	1	1.4
Total	73	

Table 3b. Sources of *lyckas* in the English-Swedish Parallel Corpus. Fiction only.

Type	Token	Per cent
Ø	39	35.5
manage	31	28.2
other	15	13.6
succeed	10	9.1
able	7	6.4
could	4	3.6
successfully	1	0.9
unable	1	0.9
get	1	0.9
do little	1	0.9
Total	110	

References

Aijmer, K. 1999. "Epistemic possibility in an English-Swedish contrastive perspective". In *Out of Corpora: Studies in Honour of Stig Johansson*, H. Hasselgård and S. Oksefjell (eds.), 301–23. Amsterdam: Rodopi.

Aijmer, K., Altenberg, B. and Johansson, M. 1996. "Text-based contrastive studies in English. Presentation of a project". In *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies in Lund, 4–5 March 1994*, K. Aijmer, B. Altenberg and M. Johansson (eds.), 73–85. Lund: Lund University Press,

Altenberg, B. and Aijmer, K. 2000. "The English-Swedish Parallel Corpus: A resource for contrastive research and translation studies." In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg im Breisgau, 1999*, C. Mair and M. Hundt (eds.), 15–33. Amsterdam and Atlanta, GA: Rodopi.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brinton, L. J. 1988. *The Development of English Aspectual Systems. Aspectualizers and Post-Verbal Particles*. Cambridge: Cambridge University Press.

Bybee, J. and Pagliuca, W. 1985. "Cross-linguistic comparison and the development of grammatical meaning". In *Historical Semantics: Historical Word-formation*, J. Fisiak (ed.), 59–83. Berlin: Mouton de Gruyter.

Bybee, J., Perkins, R. and Pagliuca, W. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Coates, J. 1983. *The Semantics of the Modal Auxiliaries*. London and Canberra: Croom Helm.

Dyvik, H. 1999. "On the complexity of translation". In *Out of Corpora: Studies in Honour of Stig Johansson*, H. Hasselgård and S. Oksefjell (eds.), 215–30. Amsterdam: Rodopi.

Edmondson, W., House J., Kasper, G. and McKeown, J. 1977. *A Pedagogic Grammar of the English Verb. A Handbook for the German Secondary Teacher of English*. Tübingen: TBL Verlag Gunter Narr.

Facchinetti, R. 2000. "*Be able to* in Present-day British English." In *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20), Freiburg im Breisgau, 1999*, C. Mair and M. Hundt (eds.), 117–30. Amsterdam and Atlanta, GA: Rodopi.

Heine, B. and Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hopper, P. and Traugott, E. C. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Klinge, A. 1993. "The English modal auxiliaries: From lexical semantics to utterance interpretation". *Journal of Linguistics* 29: 315–57.

Leech, G. and Coates, J. 1979. "Semantic indeterminacy and the modals". In *Studies in English Linguistics for Randolph Quirk*, S. Greenbaum, G. Leech and J. Svartvik (eds.), 79–90. London and New York: Longman.

Levinson, S. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Lindvall, L. 2002. "Att lyckas och att kunna. Svenska ekvivalenter till franskans *réussir* och italienskans *riuscire*". In *Festskrift till Anna Lundqvist*. Swedish Department, Göteborg University (unpublished).

Løken, B. 1997. "Expressing possibility in English and Norwegian". *ICAME Journal. Computers in English Linguistics*. 21: 43–59.

Noël, D. forthcoming. "Translations as evidence for semantics. An illustration". *Linguistics*.

Nuyts, J. 2001. *Epistemic Modality, Language, and Conceptualization. A Cognitive-Pragmatic Perspective*. Amsterdam/Philadelphia: John Benjamins.

Palmer, F. R. 1977. "Modals and actuality". *Journal of Linguistics*, 1–152.

Palmer, F. R.1986. *Mood and Modality*. Cambridge: Cambridge University Press.

Panther, K. U. and Thornburg, L. 1999. "The potentiality for actuality metonymy in English and Hungarian". In *Metonymy in Language and Thought*, K.-U. Panther and G. Radden (eds.), 333–357. Amsterdam/Philadelphia: John Benjamins.

Pustejovsky, J. 1998. *The Generative Lexicon*. Cambridge, Mass.: The MIT Press.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Stern, G. 1968 [1931]. *Meaning and Change of Meaning*. Bloomington: Indiana University Press. (Orig. published Göteborg, Elanders boktryckeri aktiebolag, 1931.)

Traugott, E. C. 1989. "On the rise of epistemic meanings in English: an example of subjectification in semantic change". *Language* 57: 33–65.

Traugott, E. C. and Dasher, R. B. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Van der Auwera, J. and Plungian, V. A. 1998. "Modality's semantic map". *Linguistic Typology* 2: 79–124.

Viberg, Å. 2002. "Polysemy and disambiguation cues across languages. The case of Swedish *få* and English *get*". In *Lexis in Contrast. Corpus-based Approaches*. B. Altenberg and S. Granger (eds.), 3–48. 119–50. Amsterdam and Philadelphia: Benjamins.

von Wright, E. H. 1951. *An Essay in Modal Logic*. Amsterdam: North Holland.

The passival and the progressive passive

A case study of layering in the English aspect and voice systems

Marianne Hundt
Ruprecht-Karls-Universität Heidelberg

Grammaticalisation of the progressive aspect and passive voice in English began in the Old English period; yet, the spread of the new patterns to all syntactic environments took several centuries. A fairly recent phenomenon is the co-occurrence of the progressive aspect and the *be*-passive. When the new form was first used in the late eighteenth century, it met with considerable resistance because it was competing with the passival, i.e. an earlier use of an active progressive with passive meaning. The study uses evidence from A Representative Corpus of Historical English Registers (ARCHER) to investigate this particular example of layering. Corpus data show that (a) the passival was not ousted by the progressive passive and that (b) this case-study of layering is a case of stable rather than transitional layering in which the older pattern (the passival) has clearly become the marked pattern.

1. Introduction

Grammaticalisation of the progressive aspect and passive voice in English began in the Old English period; yet the spread of the new patterns to all syntactic environments took several centuries. A fairly recent phenomenon is the co-occurrence of the progressive aspect and the *be*-passive (e.g. *The house is being built*). The extension of the progressive to *be*-passives appears to be a logical step in the grammaticalisation process of the progressive. However, when the new form was first used in the late eighteenth or early nineteenth century, it met with considerable resistance because it was competing with the passival, i.e. an earlier use of an active progressive with passive meaning (e.g. *The house is building*). A corpus-based study that focuses on these two competing patterns is long overdue.

A suitable source of empirical evidence for this particular example of layering in the English voice and aspect systems is ARCHER, A Representative Corpus of Historical English Registers. ARCHER spans a period of about 350 years between the end of the Early Modern Period (1650) and Present Day English (PDE). It provides data on the passival before the progressive passive was first used and allows for both patterns to be traced diachronically in the critical stages of their development. The data for both constructions were extracted from the corpus with a search for finite forms of BE followed by V*ing* (allowing for elements to occur between the auxiliary and the participle). Since the amounts of data for individual sub-periods in ARCHER are not identical, the discussion of the quantitative evidence in Section 3 will be based on normalized frequencies (per 100,000 words); absolute frequencies are given in the tables in the appendix.

As the title of the corpus indicates, ARCHER seeks to provide a representative historical sample of different texttypes or genres (newspaper prose, journals, fiction, drama, medical, scientific and legal writing, religious texts and letters). It is therefore a likely source of information on the early restrictions on the progressive passive and it might also yield evidence on the (recent) texttype-specific specialisation of the passival. Finally, ARCHER comprises samples of two major varieties of English (British and American) and is therefore expected to throw some light on possible regional differences in the survival (or revival) of the passival.[1]

ARCHER is used here to verify the functional equivalence of the passival and the progressive, and to test existing hypotheses on the origin of the progressive passive. Corpus data are also expected to clarify whether in this particular instance of layering, the parallel occurrence of the passival and the progressive passive merely represents a transition from one technique to another, or whether remnants of the passival survive in examples like *Coffee is now serving* and *The plane is now boarding*. These sentences closely resemble the passival; but rather than being a continuation of the passival, they might also be recent extensions of another pattern, the mediopassive. This construction is mainly used in the simple present or past tense (e.g. *These books sell really well*), but it is also found with the progressive (e.g. *This book is selling well*). Corpus evidence is required to verify whether the progressive passive ousted the passival, or whether examples like *Coffee is now serving* and *This book is selling well* are a direct continuation of the passival. A diachronic corpus will provide the answer to the question as to what exactly happened to the passival when the progressive passive appeared on the scene. The scenario that Bolinger (1968: 130f.) sketches in this context deserves to be quoted in length, if only for its vivid metaphors:

> Viewed close up, language changes by fits and starts. A linguistic regularity does not suddenly dissolve and reorganize itself into some new and different regularity, but fights on from the outposts even after it has lost the citadel. The seascape of our language is dotted with islands of little idioms that once upon a time embraced whole kingdoms of usage, but have shrunk to the point where all we can do is catalog them as quaint exceptions. The present passive participle vanished long ago from an expression such as *The oats are threshing*. It disappeared only yesterday from *The houses are building*, and is still part of the tricky style in *Time* magazine. It remains with us in *The coffee is making*, thanks in part to automatic percolators that have assimilated this construction to other kinds of activity that do not require the uninterrupted attention of the cook: *The water is boiling, The eggs are frying, The cereal is cooking*. The reinterpreted regularity is 'self-propelled activity', but it was not achieved overnight.

2. Theoretical background

Hopper (1991: 22ff.) introduces the term *layering* as part of a set of principles he sees at work in incipient grammaticization:[2] "The Principle of Layering refers to the prominent fact that very often more than one technique is available in a language to serve similar or even identical functions" (p. 23). These principles are not only found in grammaticization, but also in other types of change, as Hopper (1991: 32f.) points out. In a functional domain such as tense, voice or aspect, a new layer or new layers may emerge. In the functional domain of tense-marking, for instance, English has an older layer of past tense markers for verbs such as *take* (going back to the Indo-European ablaut pattern), and a newer layer of past tense marking for verbs such as *walk* (originating in the Germanic pattern of weak past tense forms). Likewise, reference to future events in English can be made through (a) the combination of *will* and the base form of a lexical verb, (b) the use of the present progressive, (c) a combination of *be* and a *to*-infinitive, (d) by using paraphrastic constructions like *be going to* or *be about to*. Layering can also be observed in the functional domain of voice at various stages in the history of English: Old English had a remnant of a synthetic passive (which had become restricted to a single verb, *hatan*), but also used infinitive constructions to express passive meaning (illustrated in example (1a)) and a combination of the verbs *beon* and *weorðan* with a past participle (illustrated in examples (1b) and (1c), respectively).[3]

(1) a. þa he þa ne mihte mid þæm hi oferswiðan, þa het he hie lædan *to beheafdianne* (*Martyrologium*, EETS, 64, 23)[4]

b. þær *wæron gehælede* þurh ða halgan femnan fela adlige menn
there were healed through the blessed woman many sick men
(*ÆLS* I 20.113)

c. þæt cweartern *wearð afylled* mid fullum adelan. and butan ælcum
the prison was/became filled with foul mud and without any
leohte atelice stincende (*ÆLS* II 35.244)[5]
light horridly stinking

The remnant of the synthetic passive, the infinitive construction and the passive with *weorðan* have all been lost,[6] but an additional layer alongside the older *be*-passive evolved in Modern English, namely the *get*-passive. As these examples show, layering can be observed synchronically, but it can also be studied diachronically. The two possible scenarios that Hopper (1991: 23) sketches for the diachrony of layering are that the old and the new form continue to co-exist (Figure 1a) or that the new form eventually replaces the older form (Figure 1b).

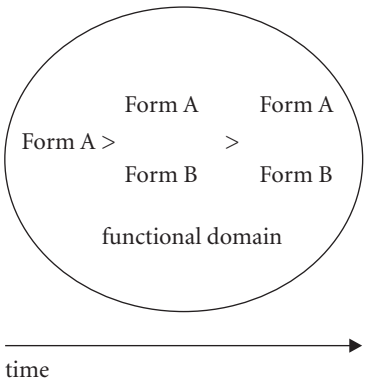


Figure 1a. Stable layering

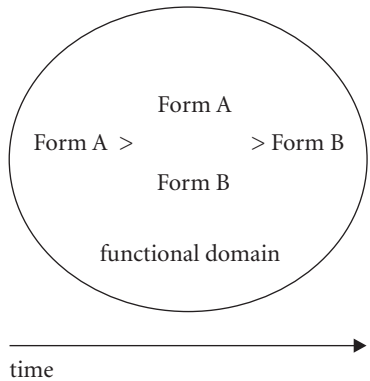


Figure 1b. Transitional layering

Note that Figure 1a does not include information on the relative frequency of the two constructions. Theoretically, both forms can be used with similar frequencies. It is more likely, though, that one form will become the unmarked pattern and the other the marked option: if layering is not merely a transitional stage, but a more or less stable phenomenon, specialisation of one form or pattern may occur, e.g. semantically, to particular lexical items or to sociolinguistic registers (Hopper 1991: 23). It is possible that the newer form remains restricted to particular usages and remains marked (see Figure 2a). Yet it is also possible that the new pattern becomes the default option and the older pattern the marked variant, probably decreasing drastically in frequency (see Figure 2b).

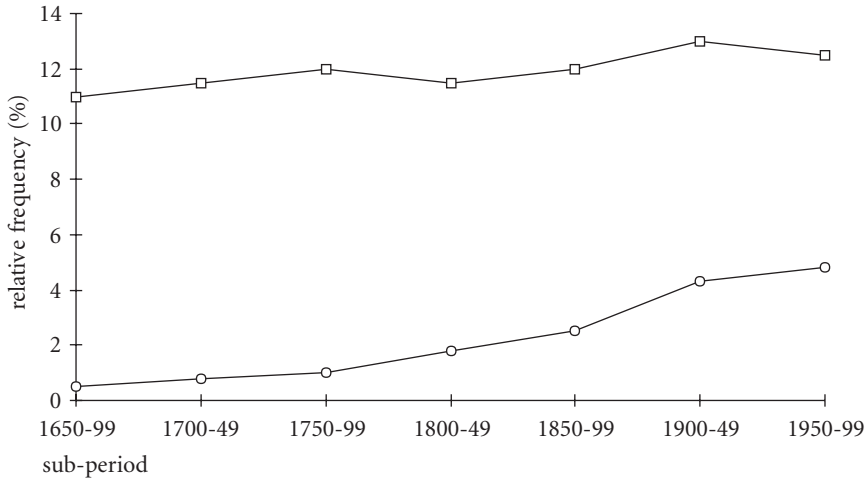


Figure 2a. Stable layering with specialisation of the newer pattern

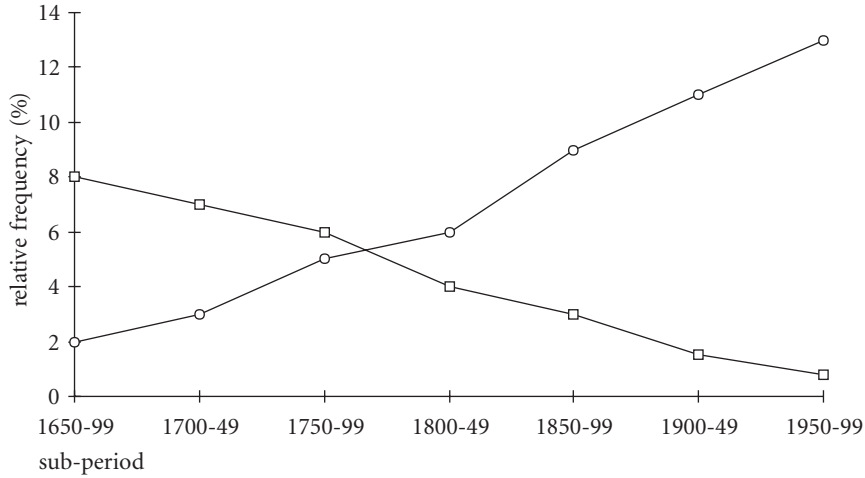


Figure 2b. Stable layering with specialisation of the older pattern

In the case of the passival and the progressive passive, the situation is slightly more complicated than in other cases of layering because the progressive passive combines two functional domains, aspect (progressive) and voice or diathesis[7] (passive). As with cases of layering in a single functional domain, layering in a construction that combines two functional domains is also expected to be diachronically either stable or transitional. Synchronically, in PDE, both progressive passives and passivals (or passival-like patterns) still

occur, as the following examples for the verb *print* from the British National Corpus (BNC) illustrate:[8]

(2) a. So I went in person to enquire what had happened to it, and after several more weeks heard that a first edition of fifty thousand copies *was being printed.* (BNC, CDC 929)
b. In the Offshore Surveys core programme, the last few sheets of the 1:250000 reconnaissance geological map series *are being printed* and published. (BNC, CFW 33)
c. A further book by Christian was published in Bombay in 1945. While it *was printing* two things happened. (BNC, CDC 1079)
d. Print speed is boosted by pages being formatted in a buffer while another *is printing.* (BNC, CP5 50)

With a verb like *print* the passival is a structural option, but *The house is building* is no longer grammatical. The question is whether the introduction of the progressive passive had a direct impact on the passival, and whether it eventually replaced the passival. Will the diachronic evidence produce a gap in the use of the passival, or are examples like *The book/page is printing* a direct continuation of the older *The house is building*?

If the introduction of the progressive passive resulted in transitional layering, the transitional phase when both constructions co-existed could either have lasted for a (short) while, or the transition could have happened more quickly, with one pattern declining as the other gained ground. The two theoretical alternatives are schematically represented in Figures 3a and 3b, respectively. (Note that both types of transitional layering are in a sense 'gradual', but that the type represented in Figure 3a has a phase of apparently stable layering — in this fictional case between 1750 and 1850 — where the substitution of one construction by the other slows down considerably or even stagnates for a while). Figure 3b looks suspiciously similar to Figure 2b. The main difference between the two scenarios is that the older variant in Figure 2b remains a low frequency (marked) option whereas it is replaced completely by the new form in the theoretical case represented in Figure 3b.

3. The passival and the progressive passive

Previous studies on the progressive passive frequently speculate on the possible origins of the construction and mention various patterns that may have

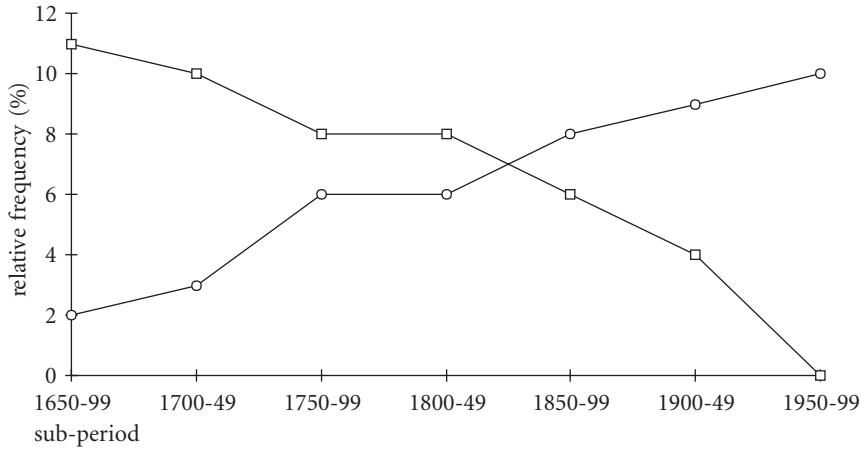


Figure 3a. Gradual transition from one pattern to another

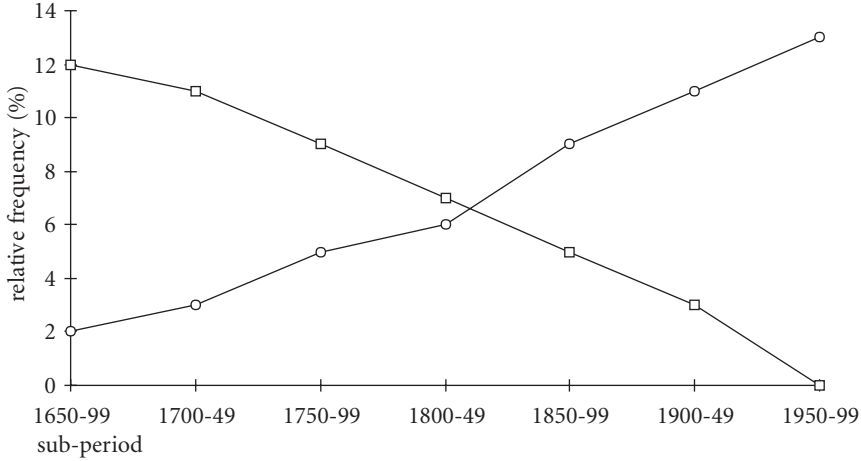


Figure 3b. Rapid transition from one pattern to another

contributed to its evolution. Quite often, linguists also comment on the fate of the passival once the progressive passive had been introduced. Both aspects of previous research provide hypotheses that can be tested with corpus data. A crucial theoretical prerequisite for the present study is the assumption that the passival and the progressive passive are functionally equivalent patterns. The following section will show whether this assumption can be sustained.

3.1 Equivalent constructions?

Hopper's (1991: 23) definition of layering is based on the assumption that the grammatical patterns involved have a similar or identical function. Yet functional equivalence of forms is very rare. Functional differentiation can even be observed in the area of past tense marking, especially when two forms of the same verb exist side by side (e.g. *learnt* vs. *learned*, *melted* vs. *molten* and *burnt* vs. *burned*). So far, the tacit assumption has been that the passival and the progressive passive are more or less equivalent patterns and that, therefore, they are an example of layering in the aspect and voice domains. Some scholars comment on differences in form and usage between the passival and progressive passive. The question is whether these differences are only minor, or whether they are so fundamental that the assumption of functional equivalence or similarity cannot be sustained. Differences between the passival and progressive passive concern the following three aspects: the expression of the agent, the semantics of the surface subject and the productivity of the two constructions. I will now look at these in turn.

– Agents in *be*-passives and passivals

Denison (1998: 149) and Visser (1963–73: §1874) claim that the agent is rarely expressed in the passival but they do not give quantitative evidence for their claim. In Visser's collection of passivals, only four contain a *by*-agent:

(3) a. Coming home ton-night, a drunken boy was carrying *by our constable* to our new pair of stocks. (1663, Pepys's *Diary*, 12 April; quoted from Visser 1963–73: §1878)

b. At the very time that this dispute was disputing *by the centinel and the drummer*, was the same point debating betwixt a trumpeter and a trumpeter's wife. (11759–67, Sterne, *Tristram Shandy*, tauchn. 23; quoted from Visser 1963–73: §1879)

c. […] it is there the search must be making *by Manfred and the strangers*. (1765 Walpole, *Castle of Otranto*, Classic Tales, 457; quoted from Visser 1963–73: §1879)

d. I remember fifty years ago, when the Battle was fighting for him *by the Few* against the Many. (1876, Letter of Edw. Fitzgerald to C. E. Norton, Wordsworth, *Letters of Ew. Fitzg.*, 1902, Vol. 2 p. 194; quoted from Visser 1963–73: §1880)

From previous corpus-based studies on the *be*-passive we know that overt *by*-agents are also rare in ordinary passive constructions (e.g. Svartvik 1966: 141

and Givón 1993: 50). Further corpus evidence is required to verify whether the incidence of *by*-agents with passivals is even lower than the reported 20% for *be*-passives. ARCHER only yields two passivals with a *by*-agent, whereas 24 of the 130 progressive passives have overt agents. At 17.7% the relative frequency of *by*-agents in progressive passives is very close to the 20% reported for *be*-passives. Even though *by*-agents in passivals are much rarer than in progressive passives, one of the two examples from ARCHER merits closer analysis:

(4) It is garded with a double mote & double wall, very strong gates, & draw bridges, we going over them were taking *by ye soldiers* & carryed before ye Mayor […]. (1687ferr.j1)

This example shares two things with the first example from Visser: it is also from the second half of the seventeenth century and, more importantly, it, too, has a human subject NP. The verbs *carry* and *take* with human subject NPs and no passive marking on the verb require a *by*-phrase if the NPs in subject position are to be interpreted as the patients. This brings us to the question of the semantics of the subject NPs in passivals and progressive passives.

– The semantics of the subject in passivals and *be*-passives

Denison (1998: 149) and Visser (1963–73: §1874) point out that the surface subjects in passivals are typically non-human or at least non-agentive. Denison (1993b: 392) mentions this as a feature that passivals share with mediopassive constructions, but prototypical *be*-passives also have affected patients as their subjects (cf. Shibatani 1985: 837 or Givón 1993: 46). The main difference between passivals and *be*-passives is that the latter involve explicit morphological marking for the passive voice on the verb and therefore are better suited for unambiguous passive meaning. While a sentence like *The bride was dressing* is ambiguous between the active and the passive interpretation, this does not hold for *The bride was being dressed.* Human patient-subjects are therefore expected to be used more frequently with the progressive passive than with the passival. In a corpus-based study, Herold (1986; quoted in Givón 1993: 69) has shown that *be*-passives in general are about as likely to occur with human as with non-human patient-subjects.

In ARCHER, the passival is used with animate subject-NPs, as the following examples from a nineteenth-century diary show:

(5) a. Countess Wedel and I remained some time in the inn while *our horses were baiting*; […]. (1827marc.j5)

b. At last, we arrived at Barum, and while *the horses were changing* we went to another manufactory to see an enormous iron hammer used for flattening iron. (1827marc.j5)
c. […] at the next post we overtook their Excellencies, who were regaling themselves with bread and cheese, while *they* [the horses, M. H.] *were changing*; […]. (1827marc.j5)

The baiting and changing of horses usually involves the agency of a human, and the passive interpretation of the examples above is therefore unproblematic. Example (4), as pointed out above, is unambiguously passive in meaning because the external agent is expressed as a prepositional *by*-agent. With the exception of example (4), the only cases of human patient-subjects in passivals from ARCHER are instances were a collective body, like an army, is involved:[9]

(6) a. […] Orders were sent into Catalonia, for 1000 Veteran Horses more to march from thence towards Madrid, whose room is to be supplied by *new Levies* that *are raising* in that Principality […]. (1704cou1.n2)
b. New *recruits are raising* to march that way. (1776king.x4)

Because the semantic frame for raising an army involves external agency, the human referents of these noun phrases are far from being volitional, controlling or actively-initiating agents. The sentences in (6) are therefore not ambiguous between an active and a passive interpretation. ARCHER does not provide evidence of potentially ambiguous sentences like *The bride was dressing*. While the passivals show a strong preference for inanimate subject NPs (30 out of 40 instances), progressive passives frequently occur with human subject NPs:

(7) a. In one *the child is being attacked* by a serpent […]. (1851carl.x6)
b. […] a *lady was being whirled along* by a couple of skaters. (1886giss.f6a)
c. Every vendable *maid* in Athens *is being walked up and down* by her mother […]. (1951andr.d0)

Of the 40 progressive passives with animate subject NPs, 39 have human subject NPs.[10] It is obvious that the progressive passive with its unambiguous passive marking is preferred over the passival with such NPs. However, of the 130 progressive passives in ARCHER, 91 (70%) have non-human subjects. More importantly, still, the first uses of the progressive passive are not instances which avoid ambiguity: their subjects are non-human NPs that could also have been used unambiguously in a passival construction. Figure 4 shows that the

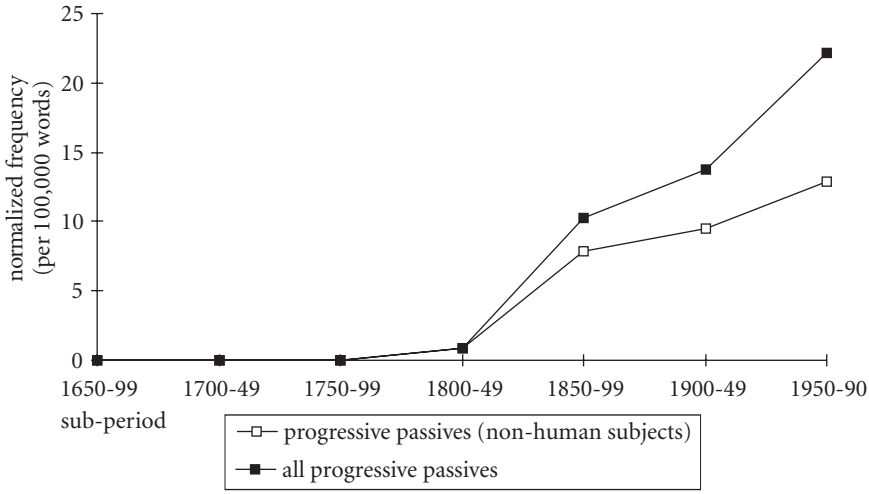


Figure 4. Semantics of the subject NP in progressive passives

steep increases in the frequency of progressive passives between 1800 and 1850 and between 1900 and 1950 can only partly be attributed to their functional potential of taking human subject NPs.

– Productivity of the passival and the progressive passive

With respect to differences in the productivity of the two patterns, both grammatical and lexical restrictions will be considered. Visser (1963–73: §1872) points out that the passival is restricted to monotransitive verbs like *build* whereas the progressive passive can also be formed from a ditransitive verb like *give*, as the examples in (8) illustrate:

(8) a. The house is building vs. *The car is giving to Henry
b. The house is being built and The car is being given to Henry

Evidence from ARCHER confirms that passivals from ditransitive verbs are ungrammatical, but it also shows that passivals from ditransitive constructions with an optional prepositional object are possible:

(9) a. A set of new ordnance, consisting of 66 pieces of brass cannon, *are preparing* at Woolwich Warren *for the new fortifications* lately raised at that place. (ARCHER, 1785gen2.n3)
b. It is said a patent of Peerage *is* now *making out for General Conway.* (ARCHER, 1773nyo2.n4)

These more unusual passivals probably find structural support in similar constructions where the prepositional phrase serves not as a complement of the verb, but as an optional adverbial:

(10) a. We hear a house in this town *is preparing for the reception of Admiral Montagu and his lady*; […]. (ARCHER, 1773nyo1.n4)
b. Arrangements are reported *to be now making for the resignation of the office of President of the Council* by the Duke of Portland. (ARCHER, 1802joh1.n5)

The second restriction on the productivity of passivals concerns the set of verbs that are found in this construction. Jespersen (1909–45, IV: 206f.), for instance, claims that from the very beginning, the passival occurs more commonly with some verbs (namely *do, make, build, print, cook* and *prepare*) than with others (see also Mossé 1938: Section 234–6). Visser (1963–73: §1879), on the other hand, includes a long list of examples for the period 1700–1799 to illustrate the productivity of the passival. The wide use of the pattern during this period is probably best illustrated with the recurrent use of the verb *to new-front* (ibid.):

(11) a. […] the house where he had his lodgings *was new-fronting* (1748, Richardson, *Clarissa* (London 1768) 3,219)
b. […] a very good house in Queen Anne style, which *is* now *new-fronting* […] (1762, Thomas Gray, *Letter to Dr. Wharton* (Everyman Library) p.251)

According to Visser (1963–73: §1881), the restriction of the pattern to certain verbs only set in with the twentieth century, i.e. at the time when the overall frequency of the pattern declined. A look at the type-token ratios in Visser's examples (Table 1) confirms this hypothesis: the variety of verbs with which the passival is used only begins to decline perceptibly in the twentieth century.

The overall frequency of the passival and the progressive passive in ARCHER is indicative of differences in the productivity of the two patterns: ARCHER yields only 40 instances of the passival and 130 instances of the

Table 1. Type-token ratios for passivals (based on Visser 1963–73: §§1879–1881)

Period	Types	Tokens	Type-token ratio
1700–1799	37	63	58.7
1800–1899	42	73	57.5
1900–	28	59	47.5

progressive passive, even though the former is attested much earlier than the latter. Visser (1963–73: §§1879–1881) has a total of 195 examples of passivals for the period between 1700 and PDE which made the calculation of type-token ratios feasible. Owing to the low overall frequency of passivals in ARCHER, it does not make sense to compare type-token ratios obtained from Visser's data with the variety of verbs found in ARCHER. What the evidence from ARCHER does confirm, however, is that the passival is used commonly with verbs like *make* (6 instances, the most common verb in Visser's eighteenth-century data) and *prepare* (5 instances, the most common verb in Visser's nineteenth-century data). *Change, fit, negotiate* and *raise* are also attested more than once, but half of the passivals in ARCHER are examples of verbs that occur only once. *Do*, the most frequent verb in Visser's twentieth-century examples, is used in only one of the passivals from ARCHER.[11] On the whole, then, ARCHER confirms that some verbs are commonly used in the passival (lexical specialisation), but the construction is not limited to only a few verbs (limited productivity).

For more information on the relative use of passival and progressive passive in PDE, I collected data from the BNC for a small set of verbs. The figures in Table 2 below are based on finite verb phrases only, excluding such occurrences as *She heard Debussy playing/being played on the piano*. Likewise, only progressive passives with inanimate subjects are included in the counts as variation between the passival and the progressive passive is very rare and sometimes impossible with animate subjects (e.g. in the case of *He was being shown around the house*).

The data from the BNC show that the progressive passive consistently outnumbers the passival. For verbs which are quite common with the passival, namely *play*, *print*, *ship* and *show*, the relative frequency of the passival amounts to figures between 28 and 36 %. The passival, while being the marked option, thus seems

Table 2. Passivals and progressive passives in the BNC

	Passival	Progressive passive	Total
DO[a]	21	360	381
PLAY	45	85	130
PRINT	8	21	29
REBUILD	1	10	11
SHIP	15	27	42
SHOW	25	63	88
SUBMIT	0	2	2

[a] Out of practical considerations, the search for this verb was limited to such patterns as *what's/what is/was doing, much doing* and *nothing doing*. The relative frequency of passivals (5.5%) is expected to be higher if all occurrences of finite BE + *doing* in the BNC were to be analysed.

to be fairly frequent with a limited set of verbs. More important seems to be the fact that the passival is not the only (idiomatic) choice with these verbs. Instead, there is variation between the progressive passive on the one hand and the passival on the other hand, as the following examples illustrate:

(12) a. William Malloch tries to prove that Haydn's and Mozart's minuets *are being played* much too slowly and ought to be rendered at considerably faster speeds. (BNC, J1 1177)
b. Thelonius Monk's ruminative 'Alone in San Francisco' *is playing* softly in the background. (BNC, FBM 710)

(13) a. Paintings of 1947 vintage by Wilfredo Lam *are being shown* throughout the month by Lelong [...]. (BNC, EBV 2694)
b. An important and unprecedented exhibition *is showing* at the Goethe Institute until 19 December. (BNC, CKY 558)

In example (12a), one could argue that the passival is not possible because the notion of musicians producing the sounds on instruments is still present; in (12b), the passival may have been used because the music is being played by a machine (a record player or a radio). In the examples under (13), however, real variation between the progressive passive and the passival is illustrated.[12] The same holds for the examples of the verb *print* under (2) above. Despite some limitations on the use of the passival, the functional equivalence between the two patterns seems close enough. They can justly be considered a case of layering.

3.2 Grammaticalisation — possible origins of the progressive passive

Various possible origins have been proposed for the progressive passive. The more recent discussions within the framework of grammaticalisation theory do not discuss this phenomenon in terms of possible grammaticalisation chains, detailing the semantic and formal changes involved in the process. Unlike the grammaticalisation of the *get*-passive and the *will*-future, for instance, the evidence we have for the grammaticalisation of the progressive passive is of a more indirect nature. It comes in the form of (a) related patterns that might have played a role in the evolution of the progressive passive, (b) patterns that evolved simultaneously and indicate the final distinction between auxiliaries and full verbs, and (c) patterns that receded as a result of the grammaticalisation of the progressive passive. Thus, these recent discussions show certain similarities with earlier work and will not be treated separately in the following overview.

Linguists like Jespersen (1909–45, IV: 210), Visser (1963–73: §2158),

Scheffer (1975: 262) and Denison (1993b: 431) assume that the gerundial and participial passive constructions paved the way for the progressive passive.[13] The two patterns are illustrated in examples (14a) to (c) and (15a) to (c), respectively:

(14) a. For, I believe he wou'd not have quarrel'd any where else, nor there neither, but upon the prospect of *being prevented*, or *parted*, *or secur'd* over night, in order to beg pardon in the morning. (ARCHER, 1692soth.d1)

b. […] scarce any Surgeon whatever can undertake an Operation of this, or indeed of any other Kind, above the Knowledge of a Child, without *being exposed*, not only to the Censures of common People, but to the stupid Restrictions […] of long-tongued, double-tongued, ignorant conceited and interested Medicaters […]. (ARCHER, 1773nyo1.n4)

c. And shall our druggist friend here insist on *being called* Doc? (ARCHER, 1970elki.f0a)

(15) a. For this *being broken* into pieces, will not so easily cement again into so compact a Body as it was formerly of, as we see in Whiting that is lighter than Chalk […]. (ARCHER, 1685slar.m1)

b. […] we must naturally infer, that woman was invested with the same, or similar privileges, and, *being* palpably *intended* for an assistant (not slave) to man, in the preservation of the species, and in procuring the necessaries of life; […] (ARCHER, 1797butl.f4a)

c. The Governor is said to have urged the expediency of avoiding the scandal which would be brought about by the cases *being heard* in open court in the presence of representatives of foreign embassies. (ARCHER, 1893man2.n6)

Scheffer (1975: 262), for instance, argues that gerundial passives started to increase in the sixteenth century, and that they occurred so frequently by the seventeenth century "that they may be said to have prepared the ground for the introduction of the passive progressive *is being built*." Figure 5 compares the development of the gerundial and participial passive constructions with that of the progressive passive.[14] ARCHER lends support to the hypothesis that the gerundial and participial passive were very frequent around the middle of the seventeenth century, but the patterns had started to decline before the first progressive passives were used. The evidence from ARCHER thus suggests that non-finite verb phrases with *being* and a past participle had occurred with a

fairly high frequency before the progressive passive was first used, but it is unlikely that these patterns were a direct precursor of the progressive passive. The connection between the patterns seems to be of a more indirect kind. The fact that gerundial and participial passives declined before the progressive passive appeared on the scene suggests that a growing reluctance to use non-finite forms of *be* in combination with a past participle may have paved the way for the progressive passive.

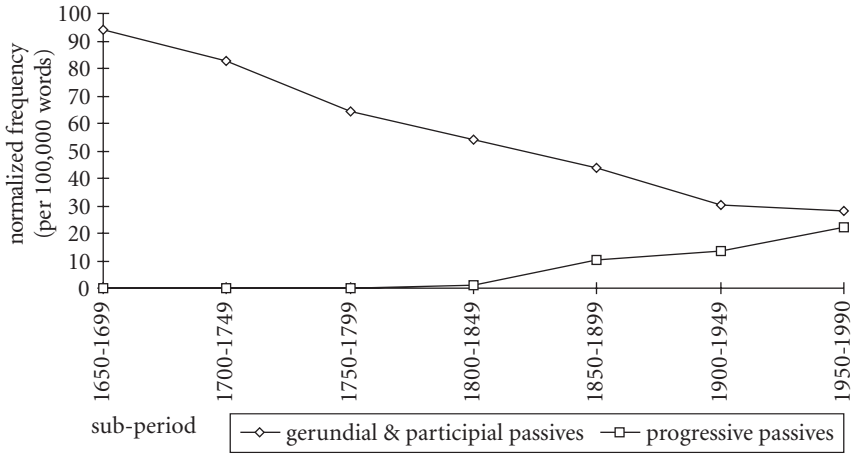


Figure 5. Gerundial and participial passives vs. progressive passives in ARCHER (BrE)

Visser (1963–73: §2158) maintains that participial constructions were only a subsidiary cause in the development of the progressive passive. Unlike Jespersen and Scheffer, he holds that the parallel (or even prior) development from *the house is to let* to *the house is to be let* was the main factor in the development of the progressive passive. A search in the first sub-period of the ARCHER corpus (BrE texts from 1650–1699), however, only yields evidence of the explicitly passive pattern (44 instances). Visser's argument also lacks plausibility in that it fails to detail how a simple infinitival pattern should have given rise to a finite pattern that is marked for aspect.

Denison (1993b: 433) suggests prepositional patterns of the type illustrated in (16a) as yet another possible precursor of the progressive passive. Example (16b), also from a journal, is the only occurrence of this pattern from ARCHER.

(16) a. He tells me that Mr. Sheply is *upon being turned* away from my Lord's family, and another sent down (1669 Pepys, *Diary* IX 475.1 (8 Mar)) (Denison 1993b: 433).

b. My Squire's man John was likely to have had a very bad accident in leading the Squire's horse over a boggy place, both horses were stuck fast up to their Bellies, and by plunging threw him of [sic] in the mire and was very *near being hurt* by the horses plunging to get out […]. (ARCHER, 1780wood.j3)

As far as discourse frequency is concerned, ARCHER does not support the hypothesis that these prepositional patterns contributed significantly to the development of the progressive passive.

Progressive *be* + adjective/NP is not a likely precursor of the progressive passive either. Most linguists agree that these patterns start occurring after the progressive passive was first used. Only Visser (1963–73: §2158) and Mossé (1938: 152) claim that the first attested examples are from the late eighteenth century, i.e. more or less contemporaneous with the first progressive passives. Denison (1993a: 23) analyses their examples as cases of copular BE and *being* + adjective; according to him, the first example of progressive BE + adjective is from 1819 (ibid.: 24). Progressive BE + NP occurs even later (Denison 1993b: 396). Jespersen (1909–45, IV: 212) points out that: "[…] one reason why people objected to *is being built* was probably that *was being* with an adjective predicative (*polite*, etc.) was not in use at that time." Data from ARCHER confirm that progressive *be* followed by an adjective or noun phrase was not a precursor of the progressive passive: patterns like *He is being silly tonight* only occurred after the progressive passive had been introduced (see Figure 6).

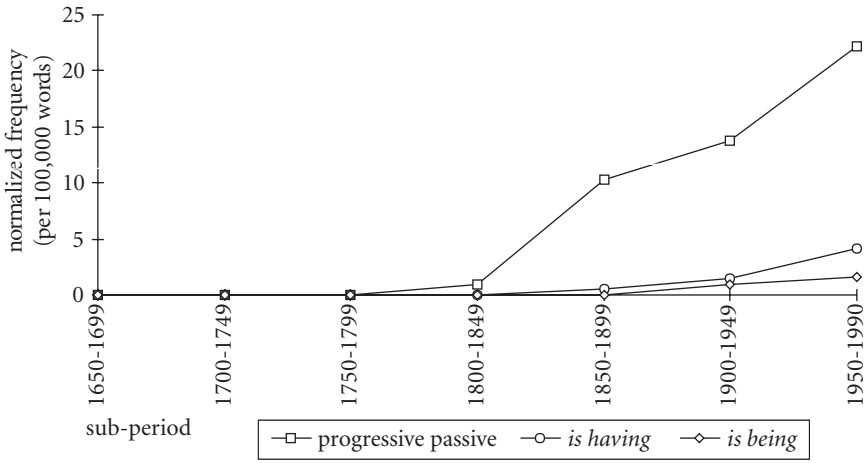


Figure 6. Progressive passive vs. progressive HAVE and BE in ARCHER (BrE)

The figures for progressive *having* and *being* include all forms (i.e. not only *is having* but also *was having*, *am having*, etc.). Absolute frequencies are given in Table 4 in the appendix.

Two patterns that are attested earlier are the combination of the participle *being* with an adjective (illustrated in (17)) and gerundial *being* + adjective, (illustrated in (18)). But like participial passives, these constructions also avoid the combination of a finite form of BE with the participle:

(17) a. Aprill ye eleventh in ye morning we came forward for Dover riding for ye most part over hills and vales, ye way *being* also *very stony*; […]. (ARCHER, 1687ferr.j1)
b. The Venetian Fleet *being too weak* to contend with the Turkish have left the important Isle of Tenedos to the Mercy of the Enemy […]. (ARCHER, 1715eve2.n2)

(18) a. My Lord, I AM to beg your Lordships pardon […] for *being silent* last week […]. (ARCHER, 1667finc.x1)
b. […] I could not reproach myself from *being too sensible* of his affection. (ARCHER, 1702anon.f2)

Denison (1993a,b; 1998) explicitly uses the framework of grammaticalisation theory to account for the fact that the grammaticalisation of the progressive passive occurs as late as it does. He argues that until the end of the eighteenth century, *being* was still used as a main verb, as for example in *Being teaching made it easier* or *Being happy is easier than being sad*. In comparison with other verbs, however, the difference between the main-verb use and the auxiliary use of *be* is less perceptible (1993b: 441), and the reanalysis from a full verb to an auxiliary therefore occurs relatively late.

> […] prior to *c.*1770 progressive BE was a main verb, from then on it could become an auxiliary, with the result that the progressive passive *was being built* was now the progressive of BUILD rather than of passive BE. The change also helps to explain the virtual disappearance at much the same time of *being Ving* […] last regularly found in Jane Austen. (Denison 1998: 155)

Denison further sees a connection between the regulation of *do* and the emergence of the progressive passive: *do* became obligatory as a dummy-auxiliary for full verbs at about the same time that the progressive of *be* grammaticalised. This resulted in an obvious difference between full verbs that needed *do*-support and auxiliary verbs (or NICE-verbs, in Denison's terminology) that did not. He speculates that this might have been "the systemic pressure

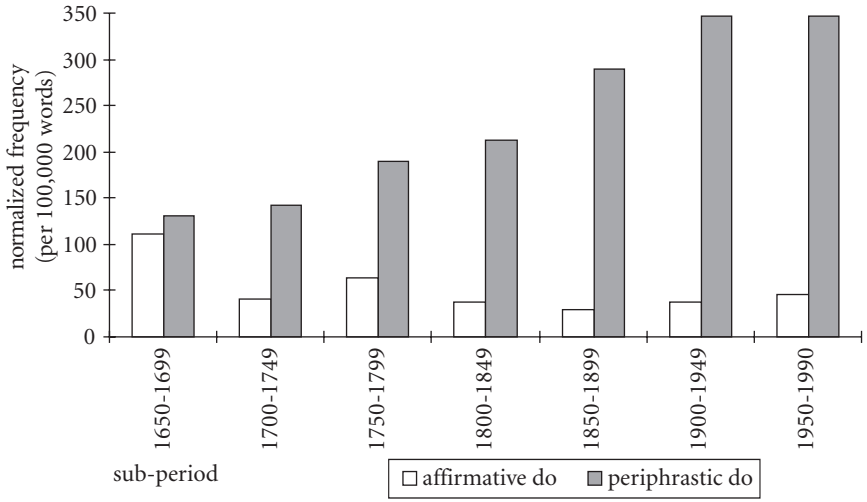


Figure 7. Periphrastic and affirmative do in ARCHER (BrE)

which brought progressive BE into line" (1993b: 443). The connection between the final regularisation of *do*-support and the emergence of progressive *be* is also made by Warner (1995), even though he sees this as a parametric change rather than grammaticalisation. The data from ARCHER, however, do not show a dramatic increase in periphrastic *do* for the sub-period 1800–1849 (the time when the first progressive passives are attested in ARCHER). *Do*-support in negative and interrogative clauses increases fairly steadily (see Figure 7).[15]

Note that for the sub-period 1750–1799, 38 of the 111 instances of affirmative *do* are from one journal. If these are omitted from the count, the normalized frequency of affirmative *do* in this period drops to 42.2/100,000 words.[16]

As part of his analysis, Warner (1995: 545) predicts that the appearance of *is having* is contemporaneous with that of *is being* as a full verb. Figure 6 shows that the first occurrence of *is having* in ARCHER is not contemporaneous with *is being*, but precedes it. Note that some earlier examples of *is being* were excluded from the data in Figure 6 because they were not instances of the full verb but the equative use of BE. In the following example, *being* is not part of the VP, but forms a constituent with *tragic*:

(19) High and tender thoughts there are, gracefully and harmoniously expressed — which is not being tragic. (ARCHER, 1840brng.x5)

A closer look at the examples confirms that progressive HAVE occurs 26 years (i.e. about a generation) earlier than progressive BE: the first attestation of progressive HAVE in the British sub-corpora of ARCHER is from 1897 (example (20a)), while the first progressive use of the full verb BE is from 1923 (example (21)). The American sub-corpus for the period from 1850–1899 provides an even earlier example of progressive HAVE from the 1860s (example (20b)).

(20) a. Her flushed silence was obvious enough for any one, except Lady Susan, who merely supposed that champagne at luncheon *was having* its almost inevitable result on the complexion. (ARCHER, 1897some.f6a)
b. The hotel here at the Springs is not so much crowded as usual and I *am having* a very comfortable time of it. (ARCHER, 1863twn.x7)

(21) If we treat others benevolently we *are* assuredly *being kind* to ourselves […]. (ARCHER, 1923step.f8)

Unlike the progressive passive, both patterns remain marginal structural options. The overall raw frequency of progressive HAVE and BE in ARCHER is very low (28 and 9 instances, respectively). The gap between the first occurrence of *is having* and *is being* in ARCHER might therefore have to be attributed to corpus size rather than historical facts. According to Denison (1993a: 24), for instance, the first example of progressive BE + adjective is from 1819, i.e. it even predates the first occurrence of *is having* in ARCHER (see also Denison 1998: 155; Nehls 1974: 160; or Warner 1995: 545).

An aspect of the final grammaticalisation of the progressive passive is its spread to more complex environments. In the twentieth century, it can combine with the perfect aspect, the future tense or a modal verb, as the following examples show:

(22) a. That er, er, little action has been taken in the last thirty forty years since this *has been being discussed*, erm, […] (BNC, JJG 541)
b. […] a rare state in our stationary process *will* just as likely *be being approached* as *being departed* from. (LOB, J18 199f.)
c. […] the seeds of any amount of trouble are sown, the harvest of which *may* still *be being reaped* at forty or fifty. (LOB, D06 17ff.)

ARCHER only provides one example of a more complex construction, the combination of the progressive passive with a non-finite form of BE:

(23) […] we were called to live and work from that for which we imagined ourselves *to be being* trained. (ARCHER, 1933hodg.h8)

Combinations of the progressive passive with the perfect aspect, the future tense or a modal verb are not attested in the ARCHER data. Such combinations are also extremely rare in PDE. Among these, the combination of the perfect with the progressive passive is the rarest construction: it does not occur in some of the standard one-million-word corpora like Brown, Frown, LOB, FLOB, the Wellington Corpus of Written New Zealand English or the Australian Corpus of English;[17] even the one-hundred-million words of the BNC only yield one example of this construction, quoted as example (22a) above. On the basis of the BNC data, Halliday's (1992: 62) impressionistic evidence seems overly optimistic, at least with respect to perfect progressive passives:

> If you listen grammatically, you will hear sentences of far greater complexity than can ever be found in writing [...]. I had heard verbal groups like *had been going to have been paying* and *will have been going to have been being tested* tripping off the tongue, at a time when structuralist grammarians were seriously wondering whether something forbiddingly complex as *has been being eaten* could ever actually be said!

Such complex constructions are possible but not widely used. Quirk et al. (1985: 167) claim that a combination of the perfect with the simple aspect is preferred over the perfect progressive passive. That passive progressives are indeed avoided in complex constructions becomes obvious once data on complex active progressives are provided: the LOB corpus, for instance, yields only two occurrences of complex passive progressive constructions, but 427 instances of complex active progressives — 167 with modals (including *will*) and 260 active perfect progressives; FLOB has only one occurrence of a complex passive progressive but 450 instances of complex active progressives — 199 with modals (including *will*) and 251 active perfect progressives.[18]

Curme (1935: 222) claims that, because the perfect and future progressive passives are too clumsy, the progressive *get*-passive is used instead. Data from ARCHER, however, confirm Denison's (1993a: 24) view that the progressive *get*-passive was not an alternative when the progressive passive was introduced. The first unambiguous progressive *get*-passives in ARCHER are from the twentieth century. Earlier combinations of BE + *getting* + V*ed* in the late 1890s are not straightforward passive constructions but patterns in which the participle could also be interpreted as having adjectival function.[19] A search in four one-million-word corpora of written standard English (LOB, FLOB, Brown and Frown) further shows that — compared with progressive *be*-passives — the progressive *get*-passive is only a marginal pattern in English. It is therefore not likely to be

Table 3. Progressive passive constructions[a]

	Brown	Frown	LOB	FLOB
GET-passive	2	2	4	4
BE-passive	171	133	196	263

[a] Instances of progressive get-passives include *getting married*. Again, figures for progressive *be*-passives in LOB and FLOB differ slightly from those given by Smith (2002).

used frequently as an alternative for perfect or future progressive passives (see Table 3). In fact, none of the examples of progressive *get*-passives in the four corpora were combinations with the perfect aspect or the future tense.[20]

Visser (1963–73: §1881) even sees the simple *get*-passive as a rival of both the passival and the progressive passive. He claims that "[…] 'the letter *gets* written', […] is expressive of the same 'aspect' as 'the letter is writing' and 'the letter is being written'" (ibid.). Historically, however, the simple *get*-passive could not have been competing with the passival either, as the former construction favours animate NPs as subjects when it is first used (see Hundt 2001) whereas the passival has a propensity for inanimate subject NPs.

Both Warner (1995) and Denison (1993a, 1998) predict that the pattern *being* V*ing* disappears with the introduction of the progressive passive (as part of the final auxiliarisation of *be*). According to Denison (1998: 204), the pattern is relatively common from the mid-sixteenth to the late eighteenth centuries.[21] ARCHER only provides few examples of this construction, but none later than the late eighteenth century:

(24) a. […] and acquainting me that he would come to me, on Monday the 13th, in the morning, before he went to Court, *being* then just *going* into the country. (ARCHER, 1749dodi.j2)
b. […] hearing of our ships standing that way, they weighed (*being watering*), in great concern, and left near 50 of their men on shore, and some water casks. (ARCHER, 1704poco.j2)
c. In the beginning of November, *being fishing* on the banks of the river Dart, which runs at the bottom of a very steep hill […] I was at once surprized with the sight of a great number of martins. (ARCHER, 1775blac.s3)

It is therefore all the more surprising that the FLOB corpus also contains an occurrence of this — nowadays apparently ungrammatical — pattern from a newspaper published in 1991:

(25) But when it comes to producing films, actress Sally Field admits: "I'd rather *being making* jam." (FLOB, C12 126ff.)

The fact that *being* V*ing* might still occasionally be used can be explained with Denison's (1998:205) words, namely that "there was no overwhelming reason" for the pattern to become ungrammatical. Occasional examples from the nineteenth and twentieth centuries can, for instance, also be found in the *OED* on CD-ROM, but *being* V*ing* is probably no more than a marginal fossilized pattern:

(26) a. […] (*a*) No special resolution..is valid..without the written consent of every person for the time *being receiving* or entitled to any relief..or other benefit from the funds of the society. (1875 *Act 38 & 39 Vict.* c. 60 §24 (8))

b. All are elders — ministers *being teaching or preaching* as well as ruling elders, and the others 'ruling elders' only. […] (1945 J. T. Cox *Practice Church of Scotland* 104)

3.3 Layering in the voice/aspect domain

Most linguists agree that the progressive passive was first used in the second half or last quarter of the eighteenth century (cf. Jespersen 1909–45, IV: 211; Denison 1993a: 22, 1993b: 428, 1998: 151; Warner 1995: 533; Bailey 1996: 222). The earliest attested written example is from 1772:

(27) I have received the speech and address of the House of Lords; probably, that of the House of Commons *was being debated* when the post went out. (1772 J. Harris, in *Ser. Lett. 1st Earl Malmesbury*, vol. I, p. 264, 8 December; first quoted in Warner 1995: 539)

Scheffer (1975: 263) claims that the progressive passive increases at the beginning of the nineteenth century and is really frequent from the second quarter of the nineteenth century (p. 263f.), a view that is also expressed by Curme (1931: 444). The passival, on the other hand, originates much earlier. According to Visser (1963–73: §1875), it dates as far back as the Old English period, but Denison (1993b: 389f.) convincingly argues that the first unproblematic instances are from Modern English. Interestingly, Visser suggests that passivals increased between 1500 and 1799, i.e. in the period leading up to the first attested occurrences of the progressive passive. Scheffer (1975: 261), on the other hand, comments on the generally low frequency of the passival: "Even when 'the house is building' was in its heyday, […] its frequency was comparatively low."

These remarks on the frequency of the two constructions are based on impressionistic evidence that needs testing against corpus data.

While there seems to be general agreement on the period when the progressive passive was first used, there is no consensus as to whether its introduction resulted in a case of stable or transitional layering. Buyssens (1979: 752), for instance, suggests that the frequency of the passival reduced when the progressive passive began to be used, but he concedes that "[…] it is necessary to measure the frequency of these two phrases in the 19th and in the 20th centuries and to compare them, which has not been done." Nakamura (1991; quoted in Denison 1998: 150), in a study based on diaries and letters, provides empirical evidence that the passival does decline sharply from the mid-nineteenth century. What he apparently did not do, however, was to compare these data with the development of the progressive passive. In a more recent paper, Warner (1995: 539) also claims that the passival was gradually replaced by the progressive passive in the nineteenth century.

Denison (1998), while acknowledging the sporadic occurrence of passivals in PDE (p. 150), postulates that the introduction of the progressive passive made the passival redundant: it lost its productivity and the remaining examples (e.g. *Dinner is cooking*) were increasingly reinterpreted as ergatives of the type *The ice is melting* (p. 155). Some of the examples Visser (1963–73: §1881) gives as twentieth-century passivals should thus be analysed as ergatives, e.g.:

(28) a. Griselda had sliced half a ham and it *was frying* on two griddles. (1933 Caldwell, *God's Little Acre* (Signet Bks.))
b. […] trying to slip into the pattern which *was* already *forming*. (1945 M. Allingham, *Coroner's Pidgin* (Penguin) 149)
c. "Hi, Patches, *what's cooking*?" (1948, J. Tey, *The Franchise Affair* (Penguin) 18)
d. […] a moment later a dozen doors *were opening and shutting*. (1950 Mervyn Peake, *Gormenghast* (London ed. 1968) 18)

Ziegeler (1999), who views the development of the progressive aspect in general as a process by which originally nominal forms obtained more verb-like properties (via adjectival and participial stages on the noun–verb continuum), gives a different reason for the disappearance of the passival. According to her, it "may be an indication of a conflict between the meanings of imperfective grammatical aspect and the perfective lexical aspectual meanings of the verbs with which they co-occurred" (p. 95). She goes on to propose that the passival

(or Middle-Voice Progressive, as she calls it) constituted an intermediate stage in the spread of the progressive to accomplishment verbs:

> The replacement of the MV-Progressive with a passive form *The house is being built* was more likely to be an indication that the increasing agentivity associated with the subject began to disallow inanimates with active verbs. (p. 95f.)

Denison (in Pratt and Denison 2000: 412) argues along similar lines: according to him, passivals ran a higher risk of being ambiguous "the more common it became for *normal* progressives to occur with non-human subjects […]."

The comments by Buyssens, Warner, Denison and Ziegeler imply that the introduction of the progressive passive resulted in transitional layering. Curme's (1931: 444) remarks, on the other hand, suggest that we might be dealing with a case of stable layering. He claims that the passival "continued to be widely used", but that this resulted in semantic specialisation. According to him, the progressive passive became the unmarked form while the passival developed semantics similar to mediopassive or ergative constructions:

> […] it came into wide use in its own distinctive field, namely, to represent an activity as proceeding easily, naturally, often almost spontaneously: 'These books *are selling out* fast.' 'Our plans *are working out* successfully.' 'Dust *is blowing in* at the door.' […][22]
>
> Thus the form with *being* and the form with the present participle were at first competing constructions without a difference of meaning, but later became differentiated, enriching the language.

Visser (1963–73: §1879–81) seems to take an intermediate position. He claims that the passival at first "seems to hold its ground firmly by the side of this innovation" (§1880) in the eighteenth and nineteenth centuries, but that it declines in frequency in the twentieth century to the point where "the construction tends to give the impression of becoming moribund" (§1881). In other words, his description appears to correspond with the development schematically represented in Figure 3a above. The view that the change from the passival to the progressive passive is likely to have been a case of gradual, rather than rapid transition is also backed by Denison (1993b: 391):

> Visser suggests that the retreat of this construction [i.e. the passival, M. H.] in the face of the advancing progressive passive did not begin until the twentieth century, which is likely enough in view of the slow spread of the progressive passive in the nineteenth.

Visser's assessment at first seems to fit in with the hypothesis of transitional

layering, but he also points out that the passival is still used with a limited number of verbs (e.g. *reprinting, doing* and *cooking*) and that these have become "permanently established idioms" (§1881). This implies that layering of the passival and the progressive passive is stable but involves lexical specialisation.

The literature on the relative development of the passival and the progressive passive does not give rise to a single hypothesis: it is not clear whether the introduction of the progressive passive resulted in stable or transitional layering. The ARCHER data (see Figure 8) show that the passival declines in frequency even before the progressive passive is first used. Empirical evidence therefore does not support Visser's claim that the passival increased between 1500 and 1799. But the introduction of the progressive passive appears to have halted the decline of the passival somewhat at first: the frequency of passivals for the sub-period 1800–1849 is even slightly higher than for the sub-period 1750–1799. Objection to the new pattern may have caused the slight increase in the use of the passival. In the second half of the nineteenth century, however, the new pattern has become more frequent than the passival ever was in the periods sampled in ARCHER.[23] Corpus data thus help us to qualify Denison's (1993b: 391) statement that the spread of the progressive passive in the nineteenth century was slow: it was slow in the first, but not in the second half of the century. The progressive passive increases from about one occurrence per 10,000 words in the first half of the century to ten in the second half.

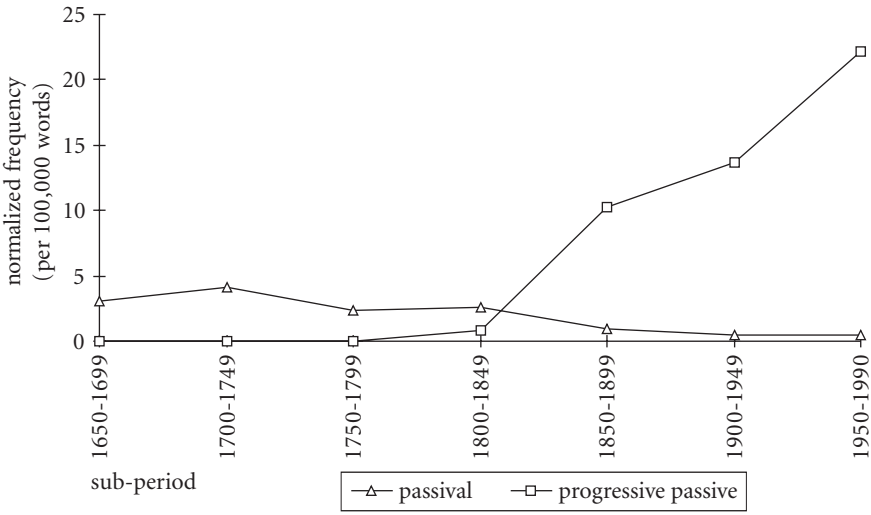


Figure 8. Passival and progressive passive in ARCHER (BrE)

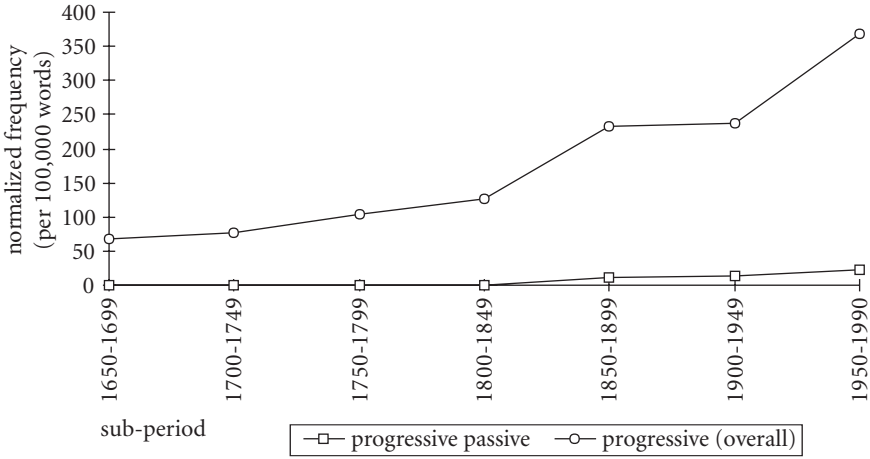


Figure 9. Progressive passives and overall development of progressives in ARCHER (BrE)

Interestingly, the greatest increase in progressive passives coincides with an even more pronounced increase in the overall frequency of progressive forms, as Figure 9 shows. In other words, increasing discourse frequency of the progressive in general supported the spread of the progressive passive. This also holds for the development in the second half of the twentieth century.

Even though the progressive passive increases steadily over time and is clearly the dominant pattern in PDE, it has not completely replaced the passival. Evidence from ARCHER thus suggests that the passival and the progressive passive are a case of stable layering. The older pattern — in this case the passival — has become the marked option over time. In which way, then, has the passival undergone specialisation?

Curme (1931:444) argues that the passival has acquired additional semantic properties which are similar to those of the mediopassive construction: according to him, the passival is nowadays used to refer to activities that proceed "easily, naturally, often almost spontaneously". Mediopassive constructions typically refer to hypothetical or potential events in which inherent properties of the subject facilitate or hinder the process expressed by the verb (e.g. *These books sell well*). Manner adverbs like *well* or *easily* are often used to focus the inherent properties of the subject in a mediopassive construction. Modal verbs (e.g. *The fax may send if you tried again*) or negation (e.g. *The fax won't send*) achieve a similar effect. The fact that mediopassive constructions typically focus on inherent properties of the NP in subject position makes them into generic

statements. In other words, mediopassive constructions do not normally refer to events in time. The typical non-eventive, generic reading is incompatible with the progressive aspect. In examples (29a) and (29b), the construction has been extended to the progressive, but these instances have to be seen as marginal members of the mediopassive construction rather than semantic specialisation of the passival:

(29) a. The 1971s, at only £1,000 a bottle, *are drinking* so much better at the moment. (*Private Eye*, 20/2/1998, p.7)
b. […] the weather was cold and windy and the course *was playing* long and tricky. (BNC, ASA 1775)

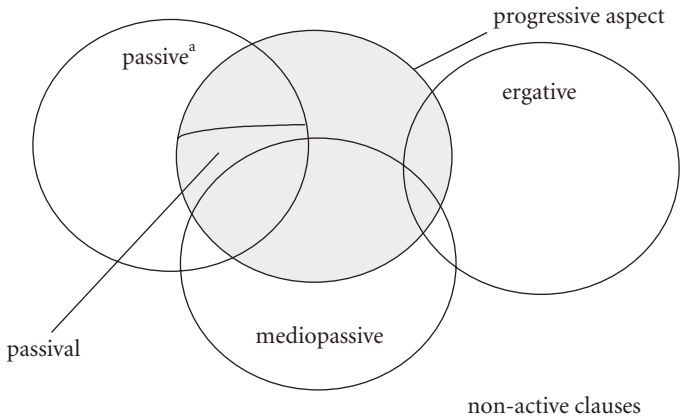
The two twentieth-century examples of passivals from ARCHER do not share the semantics of the mediopassive construction: they do not focus on properties inherent in the subject NPs that facilitate or hinder the process expressed in the verb. Instead, they are (semantic) equivalents of progressive passives:

(30) a. The book of essays *is printing* and they say they can't make alterations in the text […]. (ARCHER, 1946orwl.x8)
b. I very much doubt whether the Lyttelton will be ready for the opening season in March, which *is* now *booking*; […]. (ARCHER, 1976hal.j9)

There is thus only superficial similarity between the passival and the progressive mediopassive construction: both have non-agentive subject NPs and are marked for progressive aspect, but not for passive voice. But the two constructions do not achieve the same semantic effect. On semantic grounds, one has to argue against the hypothesis that the patterns in (29) are a continuation of the passival. Corpus evidence further indicates that passivals are continuously attested, and therefore examples like *Coffee is now serving* are unlikely to be an extension of the mediopassive construction with marking for progressive aspect — *being served* is not an inherent property of coffee whereas the drinking qualities of wine in example (29a) are inherent, if only for a limited period of time. A third pattern that is related to the progressive mediopassive and the passival is the progressive ergative construction.

(31) And in the south of England the STRUT-vowel *is fronting*, […]. (Trudgill 1998:33)

Example (31) is not a mediopassive because it does not focus on inherent properties of the subject NP; it is not a straightforward passival either, however, because the external (unexpressed) agent is not implied to the same degree as in



[a] Note that 'passive' is understood as a semantic, rather than a morphological category in this case.

Figure 10. The passival and its connection to mediopassive and ergative constructions.

the examples under (30). In other words, the process is described as occurring spontaneously. It resembles constructions of the type *Dinner is cooking*. Example (31) therefore is closer to an ergative than to a mediopassive or passival interpretation.

The connection between the examples in (29), (30) and (31) is that they are all marked for progressive aspect. In addition, they are semantically passive or passive-like, but they lack overt morphological marking for passive voice. Figure 10 tries to visualise the connection between the three patterns. Unlike the progressive mediopassive, the passival and the progressive ergative construction do not have non-progressive counterparts in the present tense: **The book prints* and **The vowel fronts*. Overlapping sections indicate that the categories are not watertight, but leak into each other, as the examples (32b) and (33b) illustrate:

(32) a. The letter you asked me to write *is printing out* right now. (passival)
b. The letter *is* not *printing out* properly somehow. (passival or mediopassive)

(33) a. The door opened. (non-progressive ergative)
b. The door opens easily. (ergative or mediopassive)

Apart from semantic and lexical specialisation, layering may also result in specialisation of one of the patterns to particular sociolinguistic registers (see Hopper 1991: 23). Comments from nineteenth-century purists who considered

the progressive passive a corrupt form of the passival imply that the new form was, at least initially, restricted to certain stylistic environments.[24] Visser (1963–73: §2158), for instance, assumes that it originated in spoken language; according to him, the fact that the first attested examples are from private letters further indicates that the progressive passive was originally avoided in writing that was to be printed (see also Scheffer 1975: 263). Denison (1993b: 29) also supposes that the progressive passive might have started out as a general low-status form that was rarely used in writing, and that later spread and became acceptable in print. He combines this hypothesis with the social-network approach, and speculates that the use of the progressive passive in writing either originated with a particular circle of acquaintances (the Southey–Coleridge circle) or was taken up by them "in a kind of inverted snobbery" (1998: 29). In a more recent paper, which takes into account earlier examples by Malmesbury (dated 1772 and 1779) and Reynolds (from 1790), Denison rejects the hypothesis that members of the Southey–Coleridge circle were the first to use the progressive passive in writing. He maintains that their adoption of the construction "at a critical period for the auxiliary system may have facilitated its later entry into the mainstream of English usage" (Pratt and Denison 2000: 419).

One of the hypotheses to be tested by the corpus-based approach is that the new progressive passive spread from speech-based genres (e.g. drama) and other informal writing (i.e. letters and journals) to more formal text types. The passival, on the other hand, is expected to show a development from a stylistically neutral to a more restricted use. Curme (1931: 444) claims that passivals increased in literary language between 1700 and 1825. While *The house is building* was first condemned as a corrupted version of *The house is abuilding* (e.g. by Samuel Johnson), nineteenth-century grammarians endorsed the passival as the correct alternative to the 'corrupt' progressive passive (cf. Visser 1963–73; Denison 1998; Bailey 1996: 223). With the growing acceptance of the progressive passive in the twentieth century, however, the tide of acceptance seems to have turned for the passival: Visser (1963–73: §1881) considers some passivals to be "slangy", and Christophersen (1952: 141) claims that the surviving passivals are either everyday phrases (like *There's nothing doing*) or patterns restricted to technical language (like *The book is printing*).

The evidence from ARCHER indicates that the passival probably never enjoyed unrestricted usage: most occurrences are from newspaper texts or journals, with only few instances occurring in more specialised registers like medical writing and similarly sparse evidence from fiction, drama and private letters (see Figure 11).

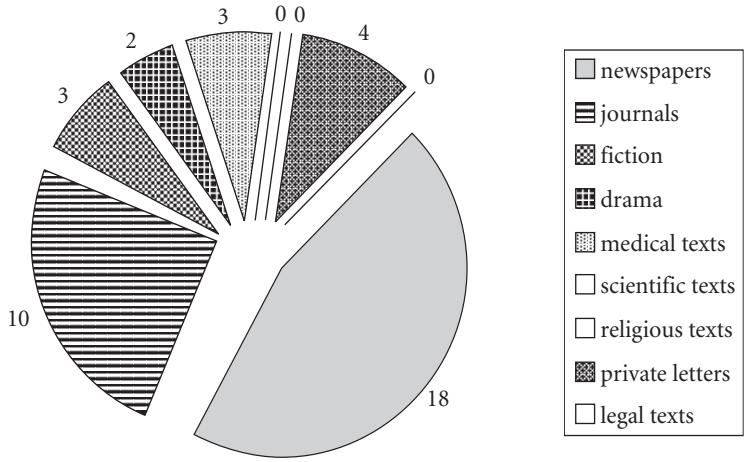


Figure 11. Distribution of passivals across texttypes in ARCHER

The overall frequency of passivals in ARCHER is too low to allow for any conclusions on diachronic developments. With only two occurrences in the twentieth century, for instance, speculations on text type-specific restrictions would be too far fetched. Data from the BNC, discussed in more detail in Section 3.1, suggest that passivals are not restricted to colloquial or technical usage in PDE. Of the 25 instances of passival *show*, 9 are from neutral reporting language in newspapers or magazines, and another 5 examples are from descriptive passages in non-fictional writing. The spontaneous conversations yield only 1 instance of this construction, and another 3 occurrences are from dialogue passages in novels. The context-governed part contains 3 stylistically unobtrusive examples of passival *show*. Overall, the 17 instances from stylistically neutral passages outnumber those from colloquial contexts.

Interestingly, the first occurrences of progressive passives in ARCHER are not from private letters but from journals and religious texts; in the next subperiod (1850–1899), they occur most frequently in newspapers, the genre that has the highest frequency of progressive passives overall (see Table 7 in the appendix). In the first half of the twentieth century, progressive passives have spread to almost all registers sampled in ARCHER (even medical and scientific writing). That progressive passives are no longer restricted to particular texttypes can also be shown with evidence from standard one-million-word corpora: in the LOB, Brown, FLOB and Frown corpora, progressive passives occur in all text categories (see Table 8 in the appendix, see also Smitterberg 2002). A look at normalized frequencies (Table 4) shows that they are marginally more

Table 4. Distribution of passive progressives across texttypes in parallel one-million word corpora (normalized frequencies per 1,000 words)[a]

	LOB	Brown	FLOB	Frown
Newspapers	0.3	0.3	0.4	0.2
Non-fiction	0.2	0.2	0.3	0.1
Fiction	0.1	0.1	0.1	0.1

[a] The newspaper section of these corpora comprises approximately 176,000 words, the non-fiction section 572,000 words and the fiction section 252,000 words. As the results based on the actual size of the newspaper sections did not differ from the approximate size, normalized frequencies are based on approximate, rather than actual numbers of words.

frequent in newspapers than in other non-fictional writing, and that they are less frequent in fictional than in non-fictional writing.

As with all linguistic patterns that were once unmarked options, it is also possible that the passival remains alive in regional varieties of English. Curme (1931: 445), for instance, claims that it is still used as the unmarked form in the southern counties of Scotland. Erades (1952: 142) also supposes that the passival is "vigorously alive dialectally." Others have claimed that *something is doing* is an Americanism (cf. Jespersen 1909–45, IV: 206).

Data from ARCHER indicate that the passival seems to have had a stronger foothold in American English (AmE) than in British English (BrE): passivals are more frequent in the transatlantic variety until the sub-period 1850–1899 (see Figures 12a and 12b). This might also explain why patterns like *What's doing* are considered to be Americanisms. Standard one-million-word corpora are too small, however, to provide conclusive evidence on regional preferences. Passival *doing*, for instance, occurs once in the Brown corpus and once in FLOB; both occurrences are from direct speech in fictional writing:

(34) a. "Nothing doing, Rufus. […]." (Brown, P10 20)
b. "Nu, nyelzya (nothing doing)," I said. (FLOB, N09 134)

While the passival had a stronger foothold in AmE, the progressive passive spread much more slowly in the transatlantic variety than in BrE. ARCHER thus supports Bailey's (1996: 222) view that the progressive passive was wrongly considered to have been introduced into BrE from AmE. Parallel corpora of Present Day BrE and AmE confirm that we are likely to be dealing with a consistent regional difference in the use of progressive passives: LOB (1961) and FLOB (1991) yield higher frequencies of progressive passives than Brown (1961) and Frown (1992), respectively (see Figure 13). It is difficult to account for this rare case of BrE being more advanced in the spread of a grammatical

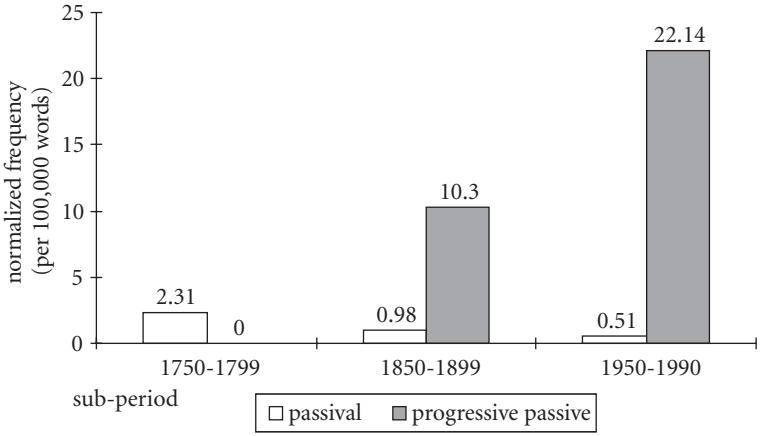


Figure 12a. Passival and progressive passive in BrE (ARCHER)

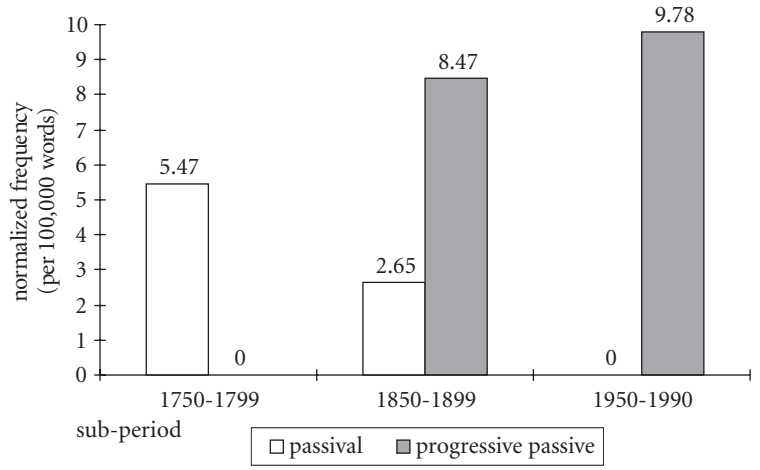


Figure 12b. Passival and progressive passive in AmE (ARCHER)

pattern than AmE. It is unlikely that the lower frequency of the construction in AmE can be attributed to prescriptive influence on the transatlantic variety: usage guides do not mention it as a problem for Present-Day AmE. According to *Webster's Dictionary of English Usage* (1989: 775), "no one thinks twice about the progressive passive", and *The Columbia Guide to Standard American English* (1993: 347) claims that "it's hard to imagine the language without it today."

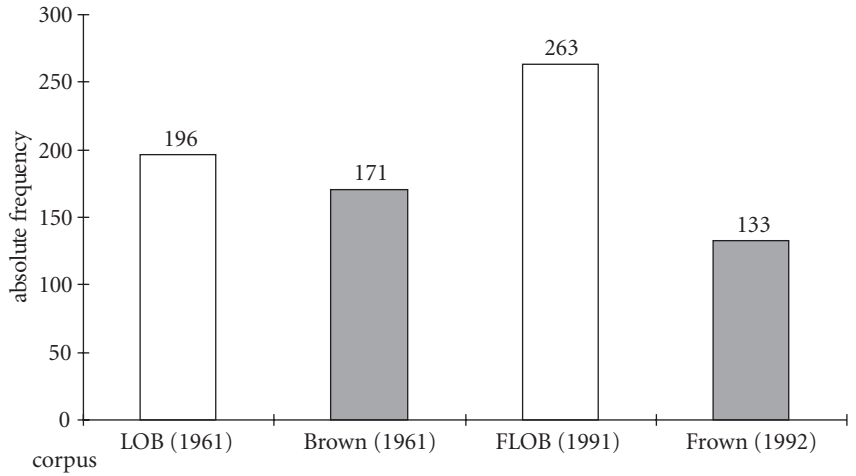


Figure 13. Progressive passives — regional differences in four parallel one-million-word corpora

4. Conclusion

The corpus-based approach taken here has provided empirical evidence that the passival was not ousted by the progressive passive. Semantic differences between the passival and the mediopassive can be used to argue against the hypothesis that the examples in (29) are a continuation of the passival. Future studies will have to show whether mixed constructions as the one in (33b) are at all frequent and might therefore be taken as counterevidence to the claim that the passival and the (progressive) mediopassive are distinct syntactic categories.

To my knowledge, the present study is the first that takes a corpus-based approach to layering as an aspect of grammaticalisation. The example studied, i.e. layering in the aspect and voice system, turns out to be a case of stable rather than transitional layering; but the older pattern (the passival) has clearly become the marked pattern. This development was gradual at first; because of the massive prescriptive reaction to the progressive passive, the introduction of the new pattern initially even halted the decline of the passival. But from the second half of the nineteenth century, the development towards near-extinction of the passival was very rapid.

Described in terms of Bolinger's (1968) geographical metaphor, even 'stable' layering can be a process in which the tide of change eats away at the cliffs of one island (the passival) and deposits the debris at the shore of another

(the progressive passive). Apart from paradigmatic pressure for the progressive passive, restrictions on the use of the passival (foremost its growing limitation to non-human subject NPs) are a good explanation as to why it lost its citadel and ceded most of its territory to the progressive passive. The question is whether the passival might eventually vanish from the seascape of our language. In other words, the final verdict might be that we have been dealing with a case of transitional layering. An argument that speaks in favour of the passival is the fact that it seems to have a fairly strong anchor in some verbs (e.g. *ship, show* and *play*). Another aspect that might help the passival to survive in its ecological niche is the fact that it is not the only construction which has passive meaning, but no morphological marking for passive voice; examples would be infinitive constructions like *The oven is still too hot to clean* or non-finite verbal complements like *This room needs cleaning*. Further structural support for the passival comes from related but slightly different constructions that are also marked for progressive aspect, but not for voice, namely examples like *Dinner is cooking* and *This wine is drinking well at the moment*. These neighbouring islands, so to speak, have provided the passival with some shelter from the currents of change. The passival is thus not a far-away, isolated island, as Bolinger claims, but part of a little archipelago.

Notes

1. For more background information on sampling procedures, sample size, etc., see Biber et al. (1994) or Biber and Finegan (1997). A new (enlarged) version of ARCHER is currently being developed in a joint project at the Universities of Arizona, Southern California, Helsinki, Uppsala and Freiburg. The data in the present paper are based on the original version of the corpus.

2. I use the term 'grammaticization' here because Hopper uses it; the term 'grammaticalisation' is used throughout the remainder of the article, but without the intention of making a theoretical point (in the way that Hopper and Traugott (1993: xv–xvi) discuss it).

3. Italics in these and the corpus examples have been added throughout.

4. Example (1a) is from Visser (1963–73: § 1873), who refers to these constructions as "passival infinitives". Note that this example has two infinitives: the first (*lædan*) is part of an ACI-construction and the second (*to beheafdianne*) is an inflected infinitive with passive meaning (i.e. 'to be beheaded').

5. Examples (1b) and (1c) are from Denison (1993b: 418, 419), who quotes them by volume, life and line.

6. PDE does not have inflected infinitives, but it uses another infinitive construction with passive meaning, e.g. in *There's more to tell* or *This oven is still too hot to clean.*

7. Strictly speaking, only the progressive passive is marked for voice whereas the passival merely has passive meaning. Some linguists (e.g. Arce-Arenales et al. 1994) reserve the term *diathesis* for the semantic concept.

8. Denison (1993b: 391) also quotes two twentieth-century examples of passivals, one with the verb *build* (from 1957) and one using the verb *set* (from 1983).

9. Jespersen (1909–45, IV:208) quotes a similar example: "[…] while the new militia was raising […]."

10. Note, however, that with only 40 out of 130 instances, the proportion of progressive passives with animate subject NPs is much lower at about 31% than the approximately 50% in Herold's study of *be*-passives (reported in Givón 1993: 69).

11. *Make, prepare* and *do* consistently rank among the four most commonly used verbs in Visser's collection of passivals.

12. The passival would be a possible alternative in (13a) (*Paintings of 1947 vintage by Wilfredo Lam are showing …*) because *Wilfredo Lam* is not the agent responsible for the showing of the paintings but the artist who produced them.

13. Note that the terminology is not always used consistently: Jespersen (1909–45, IV:210) refers to the participial passives in (15) as "gerundial passives". Visser (1963–73:§2158) does not label the participial passives but simply gives examples. A subtype of the participial passives might be patterns in which *being* is followed by an infinitival passive, as in the following example: "This day the Sheriffs […] went into the mint in Southwark to seize several Offenders, but they not *being to be found*, and no Resistance being mayde, the aforesaid Officers returned in Peace" (1682pro1.n1). This pattern is extremely rare in ARCHER, however, and all four instances are from the first three sub-periods (the latest is from 1752).

14. For absolute frequencies, see Table 2 in the appendix.

15. In a previous, corpus-based study, Nurmi (1996) focused on the connection between affirmative *do* and BE + *-ing*, but her evidence did not produce clear results supporting the hypothesis that the disappearance of affirmative *do* was connected to the appearance of BE + *-ing*.

16. For absolute frequencies, see Table 5 in the appendix.

17. Frown and FLOB are the 1990s counterparts of Brown and LOB, two one-million-word corpora of written American and British English published in 1961. For information on the 1990s corpora, see Hundt, Sand and Siemund (1998) and Hundt, Sand and Skandera (1999).

18. These figures, based on the untagged corpora, differ slightly from those obtained by Smith (2002), who used the tagged versions of LOB and FLOB.

19. Note, however, that the *OED* on CD-ROM provides an example of a progressive *get*-passive that antedates the first example in ARCHER by about 100 years: "'By Jove, I am getting hedged', thought the young man" (1863, Mrs. J. H. Riddell *World in Ch.* (1865) 66).

20. A search of *have/has/had been getting* and *will be getting* in the OED did not provide evidence of passive uses either.

21. Denison (1998:204) points out that the pattern is also found in Jane Austen's writing, but "Jane Austen is a late exponent of this usage, as she is with several grammatical features […]."

22. Note that Curme does not distinguish between examples like *These books are selling out fast* and *Dust is blowing in at the door.* The former is a mediopassive in my terminology, the latter an ergative construction.

23. In Smitterberg's (2002) data from the Corpus of Nineteenth Century English (CONCE), passive progressives also outnumber the passival by 1850.

24. For some comments, see Jespersen (1909–45, IV:212), Visser (1963–73:§2158) or Bailey (1996:222).

Appendix

Table 1. Sample size of the sub-periods in ARCHER

	Sub-period	Total number of words
1	BrE 1650–1699	162,681
2	1700–1749	170,985
3	1750–1799	173,040
4	AmE 1750–1799	164,498
5	BrE 1800–1849	230,474
6	1850–1899	203,815
7	AmE 1850–1899	189,003
8	BrE 1900–1949	211,501
9	1950–1990	194,175
10	AmE 1950–1990	194,264

Table 2. Overall frequency of progressive forms in ARCHER (forms of *going* + *to*-infinitive included)

	Sub-period	Σ	Per 10,000 words
1	BrE 1650–1699	110	67.62
2	1700–1749	132	77.20
3	1750–1799	179	103.44
4	AmE 1750–1799	127	77.20
5	BrE 1800–1849	294	127.56
6	1850–1899	474	232.56
7	AmE 1850–1899	450	250.26
8	BrE 1900–1949	611	238.09
9	1950–1990	717	369.25
10	AmE 1950–1990	674	346.95

Table 3. Gerundial and participial passives vs. progressive passives in ARCHER

	Sub-period	Gerundial and participial passives		Progressive passives	
		Σ	Per 10,000 words	Σ	Per 10,000 words
1	BrE 1650–1699	153	94.05	0	0
2	1700–1749	141	82.46	0	0
3	1750–1799	111	64.15	0	0
4	AmE 1750–1799	95	57.75	0	0
5	BrE 1800–1849	125	54.24	2	0.87
6	1850–1899	89	43.67	21	10.30
7	AmE 1850–1899	80	42.33	16	8.47
8	BrE 1900–1949	64	30.26	29	13.71
9	1950–1990	55	28.32	43	22.14
10	AmE 1950–1990	36	18.53	19	9.78

Table 4. Relative frequency of passivals, progressive passives, *is having* and *is being*

	Sub-period	Passivals		Progressive passives	
		Σ	Per 10,000 words	Σ	Per 10,000 words
1	BrE 1650–1699	5	3.07	0	0
2	1700–1749	7	4.09	0	0
3	1750–1799	4	2.31	0	0
4	AmE 1750–1799	9	5.47	0	0
5	BrE 1800–1849	6	2.60	2	0.87
6	1850–1899	2	0.98	21	10.30
7	AmE 1850–1899	5	2.65	16	8.47
8	BrE 1900–1949	1	0.47	29	13.71
9	1950–1990	1	0.51	43	22.14
10	AmE 1950–1990	0	0	19	9.78

	Sub-period	Progressive *having*		Progressive *being*	
		Σ	Per 10,000 words	Σ	Per 10,000 words
1	BrE 1650–1699	0	0	0	0
2	1700–1749	0	0	0	0
3	1750–1799	0	0	0	0
4	AmE 1750–1799	0	0	0	0
5	BrE 1800–1849	0	0	0	0
6	1850–1899	1	0.48	0	0
7	AmE 1850–1899	2	1.06	0	0
8	BrE 1900–1949	3	1.42	2	0.95
9	1950–1990	8	4.12	3	1.54
10	AmE 1950–1990	14	7.20	4	2.11

Table 5. Periphrastic and affirmative *do* (BrE)

Sub-period	Affirmative *do*		Periphrastic *do*	
	Σ	Per 10,000 words	Σ	Per 10,000 words
1650–1699	180	110.6	213	130.9
1700–1749	71	41.5	243	142.1
1750–1799	111	64.1	328	189.5
1800–1849	87	37.7	489	212.2
1850–1899	59	28.9	591	290.0
1900–1949	80	37.8	732	346.1
1950–1990	88	45.3	672	346.1

Table 6. Distribution of passivals across texttypes in ARCHER

	Sub-period	n	j	f	d	m	s	h	x	l	Total
1	BrE 1650–1699	2	3	–	–	–	–	–	–	–	5
2	1700–1749	5	–	–	–	2	–	–	–	–	7
3	1750–1799	2	1	1	–	–	–	–	–	–	4
4	AmE 1750–1799	6	–	1	–	–	–	–	2	–	9
5	BrE 1800–1849	1	3	–	–	1	–	–	1	–	6
6	1850–1899	–	1	1	–	–	–	–	–	–	2
7	AmE 1850–1899	2	1	–	2	–	–	–	–	–	5
8	BrE 1900–1949	–	–	–	–	–	–	–	1	–	1
9	1950–1990	–	1	–	–	–	–	–	–	–	1
10	AmE 1950–1990	–	–	–	–	–	–	–	–	–	0
	Total	18	10	3	2	3	0	0	4	0	

n = newspaper; j = journal; f = fiction; d = drama; m = medical texts; s = scientific texts; h = religious texts; x = private letters; l = legal texts

Table 7. Distribution of progressive passives across texttypes in ARCHER

	Sub-period	n	j	f	d	m	s	h	x	l	Total
1	BrE 1650–1699	–	–	–	–	–	–	–	–	–	0
2	1700–1749	–	–	–	–	–	–	–	–	–	0
3	1750–1799	–	–	–	–	–	–	–	–	–	0
4	AmE 1750–1799	–	–	–	–	–	–	–	–	–	0
5	BrE 1800–1849	–	1	–	–	–	–	1	–	–	2
6	1850–1899	11	–	5	–	4	–	–	1	–	21
7	AmE 1850–1899	2	–	7	–	–	–	1	–	6	16
8	BrE 1900–1949	10	4	4	2	2	2	3	2	–	29
9	1950–1990	21	3	6	1	3	4	4	1	–	43
10	AmE 1950–1990	4	5	7	1				1	1	19
	Total	48	13	29	4	9	6	9	5	7	

n = newspaper; j = journal; f = fiction; d = drama; m = medical texts; s = scientific texts; h = religious texts; x = private letters; l = legal texts

Table 8. Distribution of passive progressives across texttypes in parallel one-million word corpora (absolute frequencies)

	LOB	Brown	FLOB	Frown
A	36	24	44	16
B	16	22	27	16
C	1	8	1	5
Newspapers	53	54	72	37
D	6	3	3	7
E	27	11	28	8
F	13	13	29	9
G	24	22	35	17
H	26	21	33	9
J	21	28	26	16
Non-fiction	117	98	154	66
K	7	1	12	8
L	4	3	6	10
M	3	3	0	0
N	4	4	13	5
P	5	6	4	4
R	3	2	2	3
Fiction	26	19	37	30

References

Arce-Arenales, M., Axelrod, M. and Fox, B.A. 1994. "Active voice and middle diathesis: A cross-linguistic perspective". In *Voice. Form and Function*, B. Fox and P.J. Hopper (eds.), 1–21. Amsterdam/Philadelphia: Benjamins.

Bailey, R.W. 1996. *Nineteenth-Century English.* Ann Arbor: The University of Michigan Press.

Biber, D., Finegan, E., Atkinson, D., Beck, A., Burges, D. and Burges, J. 1994. "The design and analysis of the ARCHER Corpus. A progress report". In *Corpora across the Centuries. Proceedings of the First International Colloquium on English Diachronic Corpora*, M. Kytö, M. Rissanen and S. Wright (eds.), 3–6. Amsterdam: Rodopi. [quoted as Biber et al.]

Biber, D., Finnegan, E. and Atkinson, D. 1994. "ARCHER and its challenges: Compiling and exploring a representative corpus of historical English registers". In *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research and Computerized Corpora. Zürich 1993*, U. Fries, G. Tottie and P. Schneider (eds.), 1–13. Amsterdam: Rodopi.

Biber, D. and Finegan, E. 1997. "Diachronic relations among speech-based and written registers in English". In *To Explain the Present. Studies in the Changing English Language in Honour of Matti Rissanen*, T. Nevalainen and L. Kahlas-Tarkka (eds.), 253–275. Helsinki: Société Néophilologique.

Bolinger, D. 1968. *Aspects of Language.* New York: Harcourt, Brace & World.

Buyssens, E. 1979. "The active voice with passive meaning in Modern English". *English Studies* 60: 745–761.

Christophersen, P. 1952. "Comments". *English Studies* 33: 140–41.

The Columbia Guide to Standard American English. 1993. K.G. Wilson (ed.). New York: Columbia University Press.

Curme, G.O. 1931. *A Grammar of the English Language. Vol III. Syntax.* Boston: D.C. Heath.

Curme, G.O. 1935. *Parts of Speech and Accidence.* Boston: Heath.

Denison, D. 1993a. "Some recent changes in the English verb". In *English Diachronic Syntax*, M. Gotti (ed.), 15–33. Milano: Guerini.

Denison, D. 1993b. *English Historical Syntax: Verbal Constructions.* London: Longman.

Denison, D. 1998. "Syntax". In *Cambridge History of the English Language. Vol. IV: 1776–1997*, S. Romaine (ed.), 92–329. Cambridge: Cambridge University Press.

Denison, D. 2000. "Combining English auxiliaries." In *Pathways of Change: Grammaticalisation in English*, O. Fischer, A. Rosenbach and D. Stein (eds.), 113–47. Amsterdam and Philadelphia: John Benjamins.

Erades, P.A. 1952. "Reply to Paul Christophersen". *English Studies* 33: 141–42.

Givón, T. 1993. *English Grammar. A Function-Based Introduction.* Vol.2. Amsterdam and Philadelphia: Benjamins.

Granger, S. 1983. *The* Be + *Past Participle Construction in Spoken English with Special Emphasis on the Passive.* Amsterdam: North-Holland.

Halliday, M.A.K. 1992. "Language as system and language as instance: The corpus as a theoretical construct". In *Directions in Corpus Linguistics. Proceedings of the Nobel*

Symposium 82, Stockholm 4–8 August 1991, J. Svartvik, (ed.), 61–77. Berlin/New York: Mouton de Gruyter.

Hopper, P. J. 1991. "On some principles of grammaticization". In *Approaches to Grammaticalization. Volume I: Focus on Theoretical and Methodological Issues*, E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam and Philadelphia: John Benjamins.

Hopper, P. J. and Traugott, E. C. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Hundt, M. 2001. "What corpora tell us about the grammaticalisation of voice in *get*-constructions". *Studies in Language* 25(1): 49–88.

Hundt, M., Sand, A. and Siemund, R. (eds.). 1998. *Manual of Information to Accompany the Freiburg-LOB Corpus of British English.* Freiburg: Department of English, Freiburg University.

Hundt, M., Sand, A. and Skandera, P. (eds.). 1999. *Manual of Information to Accompany the Freiburg-Brown Corpus of American English.* Freiburg: Department of English, Freiburg University.

Jespersen, O. 1909–45. *A Modern English Grammar on Historical Principles.* London: Allen & Unwin.

Mossé, F. 1938. *Histoire de la forme périphrastique '***être** *+ participe présent' en germanique*, Vol 2, *Moyen-anglais et anglais moderne* [Collection Linguistique 43]. Paris: Klincksieck

Nehls, D. 1974. *Synchron-diachrone Untersuchungen zur Expanded Form im Englischen.* München: Hueber.

Nurmi, A. 1996. "Periphrastic DO and BE + ING: Interconnected developments?" In *Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence*, T. Nevalainen and H. Raumolin-Brunberg (eds.), 151–65. Amsterdam: Rodopi.

Pratt, L. and Denison, D. 2000. "The language of the Southey–Coleridge Circle". *Language Sciences* 22: 401–422.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language.* London: Longman.

Scheffer, J. 1975. *The Progressive in English.* Amsterdam: North-Holland.

Smith, N. 2002. "Ever moving on? The progressive in recent British English." In *New Frontiers of Corpus Research*, P. Peters, P. Collins and A. Smith (eds.), 317–30. Amsterdam: Rodopi.

Smitterberg, E. 2002. *The Progressive in 19th-century English: A Process of Integration.* PhD. dissertation. Uppsala University.

Svartvik, J. 1966. *On Voice in the English Verb.* The Hague and Paris: Mouton.

Trudgill, P. 1998. "World Englishes: Convergence or divergence?" In *The Major Varieties of English. Papers from MAVEN 97*, H. Lindquist, S. Klintborg, M. Levin and M. Estling (eds.), 29–34, Växjö: Acta Wexionensia.

Visser, F. Th. 1970. *An Historical Syntax of the English Language.* (3 Vols.) Leiden: E. J. Brill.

Warner, A. R. 1995. "Predicting the progressive passive: Parametric change within a lexicalist framework". *Language* 71(3): 533–557.

Webster's Dictionary of English Usage. 1989. Edited by E. Gilman and E. Ward. Springfield, Ma.: Merriam-Webster.

Ziegeler, D. 1999. "Agentivity and the history of the English progressive". *Transactions of the Philological Society* 97(1): 51–101.

Corpus linguistics and grammaticalisation theory

Statistics, frequencies, and beyond*

Christian Mair
University of Freiburg

The paper argues for a closer collaboration between corpus linguists and grammaticalisation theorists. Corpora have a number of benefits: First, they make it possible to study incipient or ongoing processes of grammaticalisation. Secondly, a quantitative-cum-qualitative analysis of corpus data makes it possible to shed light on important theoretical issues. Thirdly, corpora allow a better take on the textlinguistic, genre and discourse factors relevant to grammaticalisation.

The general points made are illustrated with data provided in the digitally accessible quotation base of the OED. On the basis of the evidence, I propose a distinction between two types of grammaticalisation, a "dynamic" one, which is reflected in massive shifts of frequency, and a "static" one, which can be characterised as the occasional use of lexical items in grammatical function.

1. Introduction

There is considerable common ground shared by corpus linguistics and grammaticalisation theory.

- Both approaches give priority to the study of utterances in their discourse contexts rather than abstract systems of underlying rules.
- Both emphasise the importance of frequency data and statistics.
- Both agree that transitions between grammatical categories are gradient rather than abrupt, and that grammatical form and meaning are interdependent rather than constituting separate and autonomous domains.
- Both, finally, became "hot" in linguistics again in the late nineteen seventies and early nineteen eighties after decades of relative neglect.

In spite of the common ground so obviously shared, however, corpus linguists and grammaticalisation theorists worked in blissful ignorance of each other's aims and results for quite some time. Consider, as an example, Olofsson's 1990 study of *following*, a "participle caught in the act" of becoming a preposition. It is a fine corpus-based case-study of an ongoing grammaticalisation process, but it manages not to mention the term even once. As for the grammaticalisation theorists, it is interesting to note that many of them seem to have preferred the laborious manual analysis of self-compiled and very small corpora to working with the much larger digitised databases that would in theory have been available to them (cf., e.g., Givón and Yang 1993, on the *get*-passive; Thompson and Mulac 1991, on epistemic parentheticals).

Surveying the relationship between the two fields in 1993 (published as Mair 1994), I found it necessary to trumpet the advantages to be gained from a closer collaboration between corpus linguists and grammaticalisation theorists in a paper that was largely programmatic, because there were so few examples of actual research I could have pointed out in support of my thesis. Fortunately, the situation has changed since then, and there are now a number of studies which are impressive testimony to the benefits to be derived from a closer co-operation between these two linguistic subfields. Grammaticalisation processes in early English (before 1700) have been the focus of a collection of papers produced by the Helsinki group (Rissanen et al. (eds.) 1997). A recent landmark in corpus-based grammaticalisation studies is Manfred Krug's study of emerging English modals (Krug 2000). It spans the recent history of English since the Early Modern period and is admirable in its combination of a sophisticated theoretical framework and a hands-on dexterity in the use of the relevant corpora that allows him to milk the resources to the utmost. Another recent publication relevant to the present study is Bybee and Hopper (eds.) 2001, a collection of papers devoted to the study of the role of frequency in the emergence of linguistic structure. As many contributions to that volume make clear, the role of frequency is certainly not confined to grammaticalisation alone, but — whichever way one chooses to define the phenomenon — grammaticalisation will play an important part in any such frequency-based model of grammatical structure.

Using illustrative examples from the history of English, especially the less well covered 17th to 20th centuries, the present paper will address the following three issues, which are of equal importance to the concerns of corpus linguists and grammaticalisation theorists.

1. What is the precise role of discourse frequency, and shifts in frequency, in grammaticalisation?
2. Can the use of corpora help us to pinpoint incipient or ongoing grammaticalisation?
3. What are the insights into grammaticalisation phenomena afforded by corpora that go beyond the (merely) statistical and descriptive?

With regard to the latter, more theoretical aspect, I will propose a distinction between two types of grammaticalisation: (a) a "dynamic" type which involves diachronic change and will result in drastic shifts in discourse frequencies of the constructions concerned, and (b) a "static" type, which is best described as the occasional grammatical use of lexical material. This latter type will usually not show up as a directed diachronic development in a historical corpus, nor are the phenomena concerned necessarily frequent in any statistical sense. In the main, the argument will be based on data from the *Oxford English Dictionary*. The OED's quotation base is the longest continuous historical record of any language available in digitised format, and should therefore be particularly suitable for the corpus-based study of grammaticalisation, which — as all analysts agree — is a phenomenon of *longue durée*.

2. The OED as a historical corpus

Treating an — admittedly vast — collection of dictionary citations as a corpus will come as a provocation to many who have come to see corpora as structured and balanced collections of texts compiled for linguistic analysis. The potential drawbacks of this type of "dictionary-based corpus linguistics" (Merja Kytö, personal communication) are, indeed, considerable. Rather than deal with connected passages of text, the "corpus" supplies pairs of adjacent sentences, more or less abridged sentences (the regular case), or even syntactic fragments. Many quotations turn up in several entries, and not always in identical form. Not all periods in the history of the language are covered evenly, and the editors' decisions as to what type of text should be consulted for quotations are not always in line with what today's linguist would wish for.

If the OED's quotation base is a corpus at all, it is one that rules out many types of inquiry, for example all those in which factors such as text-type specific or variety-specific frequencies play a role in the interpretation of the results. Nor is it possible to investigate macrolinguistic phenomena above the clause

Table 1. The OED quotation base — chronological breakdown

Period	Number of quotations
–1000	19,769
1001–1100	2,324
1101–1200	11,582
1201–1300	46,205
1301–1400	97,150
1401–1500	96,411
1501–1600	253,528
1601–1700	383,208
1701–1800	273,676
1801–1900	763,987
1901–2000	481,376
Total	2,428,253

level, as the textual input into the corpus is so fragmented. However, these drawbacks are offset by one crucial advantage, namely the sheer mass of material. Table 1 gives the total number of quotations per period.

The gap in the eleventh century cannot be helped; it reflects the end of the Anglo-Saxon vernacular writing tradition and/or the loss of documents. The seventeenth and nineteenth centuries are better documented than the eighteenth, but apart from the fact that this requires us to give normalised rather than absolute frequencies in comparisons, the mild ups and downs in coverage are not really serious. The nineteenth and twentieth centuries are represented particularly well. If one assumes an average length of ten words per citation — a very conservative estimate —, we have 7.64 million words for the nineteenth, and 4.81 for the twentieth century. This is much more than is provided for the respective periods in carefully compiled corpora such as ARCHER ("A Representative Corpus of Historical English Registers"). As for the editors' "Victorian" bias towards the British, the literary and the polite ends of the textual spectrum, much has been redressed in the second edition. Four letter words and Americanisms are now in — just to mention two oft-deplored gaps.[1]

The idea of using the OED's quotation base as a historical corpus grew out of our Freiburg-based project on the corpus-based investigation of change in present-day standard English.[2] The 30-year reference interval separating our set of British and American matching corpora is insufficient to establish whether a frequency shift observed in our material might plausibly be regarded as an episode in a longer process of grammaticalisation. As one of the project's more

ambitious goals is to uncover changes not suspected or described in the linguistic literature, the time horizon had to be pushed back in such cases of suspected grammaticalisation. ARCHER (present-day English to 1650) and the Helsinki corpus (Early Modern English to Old English) provide a continuous historical record, but very often neither corpus contains sufficient amounts of material for analysis. This is why we often ended up with the OED as a last resort. Our experience suggests that if the focus of the investigation is on clause-level grammatical phenomena, and the level of generalisation aimed for is the development of the English language in very general terms rather than the history of any one specific variety, style or register, then the advantages of this database far outweigh the drawbacks enumerated above.

3. Grammaticalisation and frequency

Grammaticalisation is tied in with frequency from the very outset, as it were. The starting points for most grammaticalisation processes are general-purpose words with rather broad meanings (e.g. *go* rather than *walk, get* rather than *receive*); these may enter into a broad range of possible combinations with other lexical items and are usually rather frequent. In addition, however, discourse frequency is met with as an important factor in the discussion of many specific aspects of grammaticalisation.

Increases in the discourse frequency of lexical items may, of course, have many different causes — many of them completely unrelated to grammaticalisation. For example, the growing "visibility" of women in the public sphere in Western industrialised societies in the past few decades has led to a higher incidence of feminine pronouns in recent written corpora. Certain types of semantic change, such as, for example, the extension of the meaning of a word, will show up in increased discourse frequency — whether the development is part of a grammaticalisation process or not. Such caveats notwithstanding, though, it is reasonable to assume that, given the right context, sudden and otherwise-unaccounted-for increases in discourse frequency may be regarded as symptoms of incipient grammaticalisation. For processes of grammaticalisation that are already on the way, discourse frequency is expected to rise in proportion to the increasingly grammatical nature of a word, morpheme or construction.

Hopper and Traugott, the authors of an important textbook on grammaticalisation, for instance, emphasise the importance of studying "patterns of

usage, as reflected by the frequency with which tokens of these structures may occur across time" (1993: 59). Very much in the spirit of sociolinguistic and corpus-based approaches to diachronic change, they insist that, like any other kind of change, grammaticalisation is embedded in variation:

> Typically […] the initial stage is already one of variation, and the final exemplified stage may still be in variation. Such quantitative studies highlight the gradualness of the spread of changes. (1993: 60)

Where recent approaches to syntactic change emphasise the long temporal sweep and wholesale and systematic re-setting of parameters (Lightfoot 1979 to 1999), Hopper and Traugott again echo corpus-linguists in their insistence that the data are messier than most generalisations:

> A particular grammaticalization process may be, and often is, arrested before it is fully 'implemented', and the 'outcome' of grammaticalization is quite often a ragged and incomplete subsystem that is not evidently moving in some identifiable direction. (1993: 94)

The importance of frequency is elevated to a level even a corpus-linguistic number-cruncher might envy in remarks on the relative degrees of grammaticalisation of vector verbs in Hindi-Urdu and Kashmiri: "The sheer textual frequency is prima facie evidence of degree of grammaticalization." (1993: 110) They conclude the discussion with a welcome general plea: "There is an urgent need for additional reliable statistical studies of a variety of phenomena in which early grammaticalization appears to be involved." (1993: 112)

This optimism notwithstanding, however, I would argue that we are still far from a precise understanding of the role of frequency in grammaticalisation. In the published literature, for example, there is little to guide us on the most crucial questions: Is an increase in discourse frequency a prerequisite for and concomitant of ongoing grammaticalisation (the view one would come away with from reading Krug 2000), or is it a mere epiphenomenon, a post-facto symptom that grammaticalisation has occurred and the newly developed structural option is spreading through genres and styles (as is suggested by the findings of Hundt 2001)? Nor do we know much about the relationship between discourse frequency and psychological salience for speakers and listeners, who after all are the human agents bringing about the changes that reveal themselves as grammaticalisation phenomena in retrospect.

In order to shed light on such issues, I will take Hopper and Traugott's own major study example, the English *going to*-future, and compare their account with the observable increase in frequency in the OED quotation base. Further on,

I will turn to several more problematical cases, which involve incipient, incomplete, or otherwise complicated grammaticalisation:

1. the use of *begin* and *start* as aspectual verbs, that is semi-auxiliaries complemented by either an infinitive or an *-ing* complement;
2. the increasing use of bare infinitives with *help* — not a phenomenon usually discussed as such in the literature on grammaticalisation but nevertheless showing striking parallels (as I have demonstrated elsewhere — Mair 1995);
3. the verb *see*, which is an extremely frequent and extremely polysemous lexical verb which has shown signs of grammaticalising into several directions, for example by giving rise to deverbal conjunctions (cf., e.g., *seeing that the sum total is different from the previous one, there must have been an error* or *she looked to see who was approaching*), or being used in diathesis[3] (cf., e.g., *I'd hate to see you lose* vs. *I'd hate you to lose* or *the polls see Gore in the lead* vs. *we see Gore leading in the polls*);
4. the verb *suppose*, which is the common source of two different instances of grammaticalisation, namely the modal idiom *be supposed to* and the pseudo-conjunction *supposing (that)*.

In terms of the twofold distinction proposed above, the aspectuals, grammaticalising *help*, *be supposed to* (and, of course, *going to* itself) represent the dynamic type of grammaticalisation, with clearly directed diachronic changes and associated shifts in discourse frequency. The complex conjunctions (*seeing (that), supposing (that)*), on the other hand, represent the static type, with no clear diachronic trends or shifts in frequency observable in the data.

4. The delayed increase of discourse frequency in mature grammaticalisation: The case of *going to*

While, strictly speaking, there is no precise end-point to a grammaticalisation process — after all, phonetic reduction and morphological incorporation can always proceed further after the most important semantic and syntactic settings have been decisively and irreversibly switched —, most research agrees that the grammaticalisation of *be going to* from "progressive of a motion verb followed by infinitive of purpose" to "indicator of future" has been completed. This makes it particularly interesting to bring together the standard analyses of the process as understood and described in the language-historical literature, including

Hopper and Traugott's textbook, and the frequency data in the OED corpus.

According to Jespersen, the grammaticalised use "began towards the end of the 15th c., but is not yet frequent ab[out] 1600" (1909–1949: IV, 217). The relevant OED entry (47b) gives a first good example for the year 1482[4] and provides continuous documentation from the late 17th century onwards. Joshua Poole's *English Accidence*, published in 1646, explicitly recognises *going to* as a future marker, which strongly, if indirectly, suggests that grammaticalisation was well under way by that time (see Danchev and Kytö 1994: 67 for a quotation and discussion of the relevant passage of the work).

Quantitative data is available on the subsequent spread of the *going to* future, usually obtained by calculating the proportion of *going to*-futures and *shall/will*-futures in a given work of literature (e.g., Visser 1973; cf. also the review of similar work in Danchev and Kytö 2002). In addition, there are specialist corpus-linguistic studies of *going to* — in particular Danchev and Kytö 1994 for the earliest period (based on the Helsinki Corpus) and Mair 1997 for the very recent past (based on Brown, LOB and their Freiburg updates Frown and F-LOB). However, as the corpora sampled are either too small or too heterogeneous to allow easy comparison, no coherent historical record emerges from these studies. It is here that the analysis of the OED quotation base comes in useful.

As the following graph, indicating the occurrence of *going to* per 10,000 quotations, shows, a marked rise in frequency did not occur until the end of the 19th century, but has continued unabated since then.

In order not to be misled by purely quantitative measures, I singled out the attestations from the last quarter of each of the four centuries surveyed for close analysis, separating instances of prepositional and infinitival *to*. This yielded the following, somewhat differentiated picture.

On the language-historical record, grammaticalisation was complete by the end of the 17th century. This was at a time that (see Fig. 1) the increase in frequency had not even got under way. But it was also the time at which

Table 2. *Going to* — manually post-edited output for four quarter centuries

	Going **+ infinitive (absolute frequencies)**	***Going*** **+ infinitive (as n/10000 quotes)**	***Going*** **+ infinitive (as per cent of all instances)**
1676–1700	20	2.5	51.3
1776–1800	37	4.7	59.7
1870–1900	234	9.1	76.5
1976–2000	158	34.2	90.8

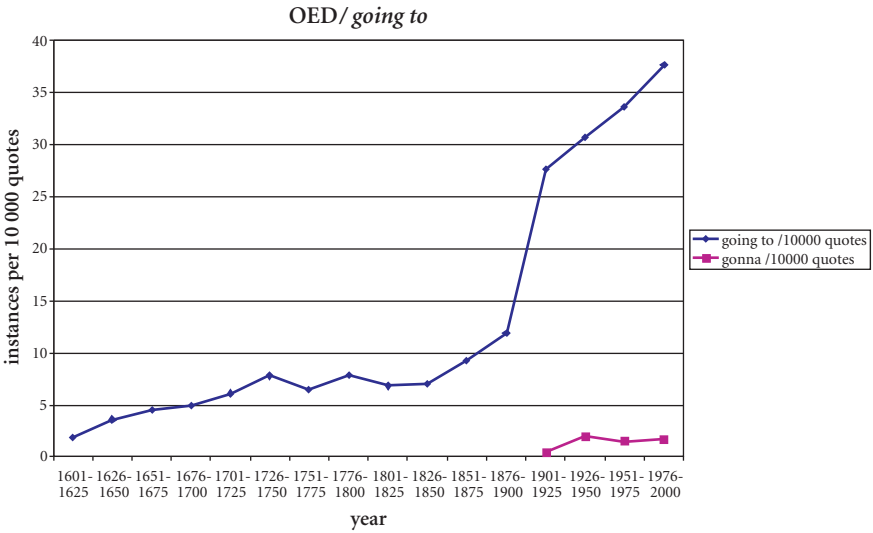


Figure 1. *Going to* and *gonna* 1600–2000 — frequency as *n*/10,000 citations

infinitival *to*, the relevant environment for grammaticalisation, already accounted for more than half of all instances of *going to*. It may well be this latter statistical fact which is the crucial indicator of grammaticalisation occurring (or having occurred), while the subsequent rise to near saturation point (90+ per cent for the present), accompanied by the drastic increase in the over-all absolute frequency of *going to*, documents the relatively rapid spread of a successful innovation through different styles and genres.[5] It will, of course, be interesting to see whether the statistical facts reported here are unique to the case of *going to*, or whether comparable frequency patterns can be observed in other grammaticalisation processes.

5. Is the delayed-increase pattern generalisable? The case of *begin/start*

Up to the 18th century, *start* was a verb with a range of rather specific lexical meanings, much like its German cognate *stürzen* is today. By 1800, however, its meaning had widened to include a generalised sense of inception, in which it grammaticalised rapidly as an aspectual semi-auxiliary that took a place alongside *begin* (which had been used in this function for the whole recorded history of English). As Figure 2 shows, the quantitative patterns for *start* (followed by an *-ing*-complement or an infinitive) closely parallel those

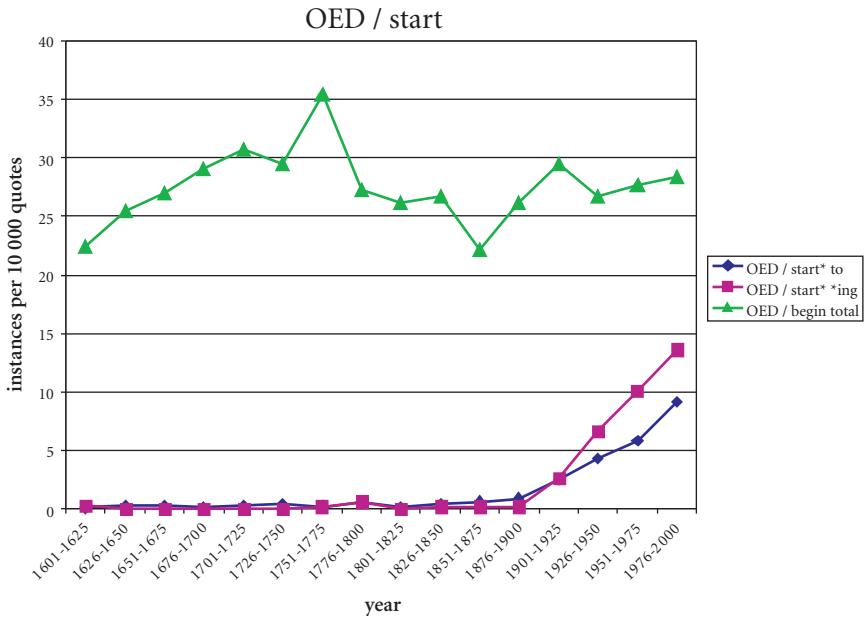


Figure 2. *Begin/start* + gerund and infinitive 1600–2000 – frequency as n/10,000 citations

obtained for *going to*. That is, the frequency of aspectual *start* remains low throughout the 19th century, and only then do we have the expected dramatic rise in discourse frequency. Another thing the graphs show is that the spread of *start* as an aspectual semi-auxiliary did not proceed at the expense of *begin*, so that the development has been part of a strengthening of the category of aspectuals as a whole (cf. also Brinton 1988).

Again, it is instructive to complement the quantitative analysis with qualitative follow-up studies of parts of the material.

A first and somewhat disappointing result to be gleaned from the table is the unreliability of the unedited figures (visualised in Figure 2 above and reported in the Appendix), especially in the 18th and 19th centuries. Unlike *going*, which is morphologically identified as a participle, *start* is not specified with regard to part of speech or grammatical function. This means that the search for *start* to*, while capturing all the relevant forms, overcollected many instances of the noun *start* followed by a preposition, for example, the common idiom *from start to finish*. What Table 3 shows beyond doubt, however, is that — as in the case of the *going to*-future — there is a delay between the crucial stage in the grammaticalisation of *start* as an aspectual verb (presumably late in

Table 3. *Start* — manually post-edited output for six quarter centuries

	All verbal uses	Aspectual use + infinitive	Aspectual use + gerund	Aspectual uses as per cent of all verbal uses
1726–1750	14	–	–	0
1776–1800	31	–	–	0
1826–1850	74	1	–	1.4
1876–1900	191	7	1	4.2
1926–1950	219	16	36	23.7
1976–2000	78	9	9	23.1

the 18th or early in the 19th century) and the rise in frequency (which is essentially a twentieth-century phenomenon). On the other hand, it is interesting to note that grammaticalised uses seem to have levelled off at around a quarter of all verbal uses of *start* — which is a long way from the 90% observed for *going to* and may reflect a weaker degree of grammaticalisation.

6. The delayed-increase pattern as a discovery procedure in cases of ongoing grammaticalisation: The case of *help*

Whether it is possible to recognise early grammaticalisation is a point considered controversial in the literature. Among the seven "themes of current concern" listed in Hopper and Traugott (1993: 31), the question "when can incipient grammaticalization be recognized?" ranks fifth, and some of the remarks quoted in the preceding section of the paper suggest that for these two authors shifts in textual frequency are important indicators of early grammaticalisation.

Other writers, by contrast, are extremely sceptical about the possibility of observing grammaticalisation processes unfolding in the field:

> Die früheste Phase der Grammatikalisierung — also die Entstehung von Formeln und einfachen 'pattern' — aktuell zu verfolgen, ist phänomenologisch unmöglich. Ungeheure Massen an Diskursen mit einem potentiell grammatikalisierenden Kandidaten müßten aufgezeichnet und, vor allem ausgewertet, werden, doch nach welchen Kriterien? Und welche sind die Kandidaten? (Compes, Kutscher and Rudorf 1993: 20)[6]

Christian Lehmann seems to take a middle road in his study of ongoing change in present-day German, which ends on the following, decidedly sceptical note:

> The study of grammatical change on the basis of synchronic (or, at best, brachychronic) variation in the contemporary stage of a language is subject to a serious problem of verification. Given presently available methodological means, it is next to impossible to know which of the changes that speech habits currently exhibit are synchronic manifestations of ongoing language change, and which of them are but ephemeral fashions. In this situation, a study such as this can hope to elucidate the synchronic dynamism of the language; it cannot hope to tell which innovations will result in changes and which ones will disappear without a trace in future synchronies. (1991:532)

Without going into detail (which I have done elsewhere — cf. Mair 1994) I would argue that the growing body of diversified corpus-linguistic resources available for a language such as English would go a considerable way towards remedying the problems pinpointed by Compes et al. and Lehmann.

In the context of the Freiburg project's narrow-scope study of changes in standard English between 1961 and 1991/92, I noted a fairly dramatic increase in the frequency of *help*, especially when the verb was followed by a bare infinitive, which I provisionally interpreted as ongoing grammaticalisation (Mair 1995). *Help*, I argued, was losing its original lexical meaning and turning into either a deverbal preposition (cf., e.g., (1) and (2) below) or a catenative verb in a catenative+lexical verb construction of the type so common in Modern English (cf. (3)):

(1) To help pay for the national debt, Congress voted to increase taxes.

(2) To help fight mad-cow disease, it is necessary to impose controls on farmers.

(3) Will tougher laws help stop street-crime?

Given that a period of thirty years is an extremely short interval in a typical process of grammaticalisation this analysis remained speculative. It is, however, corroborated by the long-term trends as documented in the OED quotation base.

If we assume rough parallelism between Figures 2 and 3, then the origins of the grammaticalisation process should be located in the late 18th century, when they went unrecognised by contemporary observers. A striking sign of this is the relevant entry in the first edition of the OED, in which the bare infinitive after *help* was considered a dialectal or vulgar form — an error duly corrected in the second edition published in 1989. Numerous sources, from early sources through Quirk et al. (1985:1205) up to Biber at al. (1999:735–737), though not the recent Huddleston and Pullum (2002:1229–30, 1244), misinterpreted *help* + bare infinitive as a syntactic Americanism, which is true merely in the

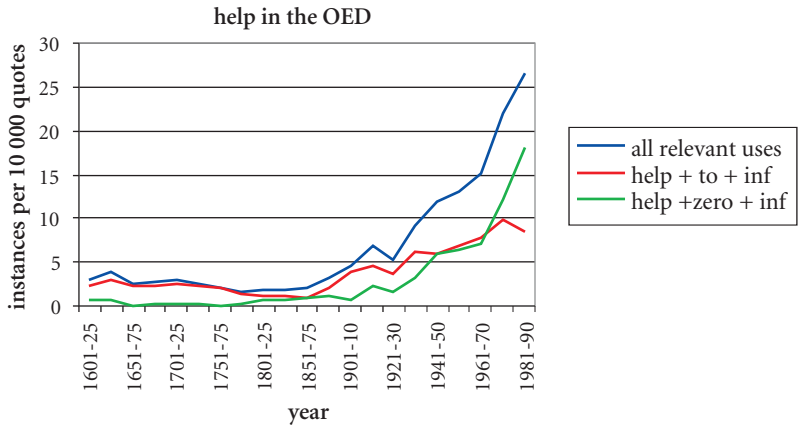


Figure 3. *Help* + infinitive 1600–2000 — frequency as *n*/10,000 citations

tenuous sense that the grammaticalisation process may for some time have proceeded faster in American English than in the British variety.

In terms of methodology, it must be noted that, with *going to* and *start*, we assumed grammaticalisation on independent grounds, and the results from the OED merely served to corroborate a pre-existing plausible analysis. With *help*, however, it is the corpus findings themselves which have alerted us to the possibility of such an analysis in the first place — thus showing that the identification of statistical patterns commonly associated with known cases of grammaticalisation may serve as a discovery procedure in the search for similar, but hitherto unrecognised instances.

7. Grammaticalisation against a background of statistical noise: The case of *see*

As already hinted at above, not all grammaticalisation phenomena leave a statistical imprint in diachronic corpora. As Figure 4 shows, *see* is a good case in point.

There is considerable, probably random, variation here, but with no perceptible trend. What is worse, the participle *seeing*, which — as the basis of the deverbal conjunction *seeing (that)* — is most specifically relevant to grammaticalisation, has remained constant in its frequency. There are several possible ways of accounting for this unexpected result:

1. Lexical uses of *see* are so frequent that the development of the grammatical ones is rendered "invisible" in the statistics. That is, statistical patterns

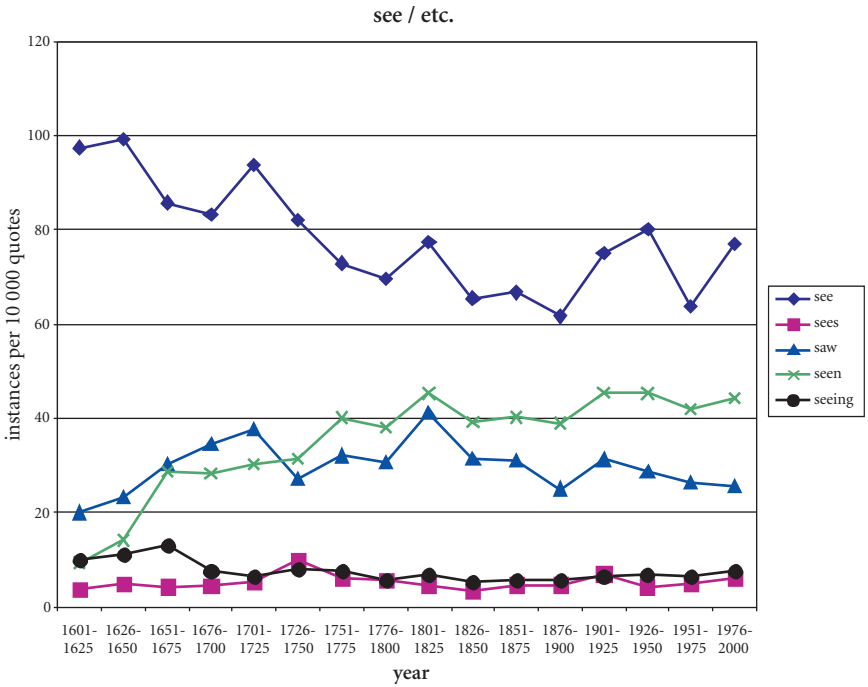


Figure 4. *See** 1600–2000 — frequency as *n*/10,000 citations

similar to those observed in the cases discussed above are present, but they are drowned out, as it were, by statistical noise.

2. Grammaticalisation has proceeded, but without the expected shifts in frequency.
3. Grammaticalisation as evident in late Middle English has not progressed for the past 500 years.

Faced with a choice between these alternatives, there is no solution but to give up the superficial but time-efficient statistical analysis of the data in favour of the traditional philological methods with their emphasis on the close reading of individual instances in their textual context. This is not only time-consuming but additionally problematical because of the relative decontextualisation of the dictionary citations, which frequently makes them difficult to interpret and categorise unambiguously.

Such problems notwithstanding, however, a closer look at attested examples suggests that it is hypothesis (3) which is the correct one. Precisely the type of mix between clearly lexical, vague or ambiguous, and clearly grammatical uses

of *seeing (that)* which characterises present usage already existed in the Early Modern period, as a brief collection of examples will illustrate:

(a) *seeing (that)*: clearly lexical uses

(4) Leosthenes seeing that he could not by force winne the towne, straightwaies cut of their victuals. (1569 T. Stocker tr. Diod. Sic. i. iv. 9)

(5) Thieriot⟨dd⟩seeing⟨dd⟩ that the wind was now blowing in Voltaire's direction, consented⟨dd⟩to give the required evidence. (1957 N. Mitford Voltaire in Love x. 115)

(b) *seeing (that)*: indeterminate between lexical and grammatical readings

(6) Why haue I found grace in thine eyes, that thou shouldest take knowledge of me, seeing I am a stranger? (1611 Bible Ruth ii. 10)

(7) The pure abstractionist will come forward with a question: Seeing that the non-representational passages in representational works are so expressive, why should anyone bother with representation? (1950 A. Huxley Themes and Variations iv. 184)

(c) *seeing (that)*: clearly grammatical uses

(8) Seeing that the King could not be reformed by Sute of Law, that ought to be done by aspertee that is by force. (1660 Bond Scutum Reg. 234)

(9) And seeing we haue you here alone, your stearne lookes shall stande for no sterling. (1584 Greene Mirror Modestie Wks. (Grosart) III. 25)

(10) The 'Wide Screen' invention, though perfected, was not offered to the public by the big producing concerns, seeing that it would involve the studios in huge expenditure. (1932 Ibid [= *Ann Reg*] 1931 47)

Can we have grammaticalisation with nothing happening for four hundred years? Not if we insist that all grammaticalisation phenomena need to be instances of the "dynamic" type described above.[7] However, as Traugott and Heine have noted:

> There is more significant disagreement about whether grammaticalization is primarily a diachronic phenomenon to be studied from a "source and pathway" perspective, or primarily a syntactic, discourse-pragmatic phenomenon, to be studied from the point of view of fluid patterns of language use across time or at a synchronically segmented moment in time. (1991: 1)

On the latter view, and for perfectly good reasons, grammaticalisation could be argued to include not only directed diachronic processes but also those cases in

which a latent (and ever-present) option is exercised to use suitable lexical items in chiefly grammatical functions. This is the second, "static" type of grammaticalisation which is not associated with language change but manifests itself in a diachronically stable "corona" of marginal and experimental uses around some highly frequent lexical items. Since they piggyback on suitable lexical items, such uses can be entrenched without being frequent.

8. Quantitative and qualitative evidence: The differential grammaticalisation of *suppose*

Like *see*, the lexical verb *suppose* is the source of at least two different instances of grammaticalisation.

a. the modal idiom *be supposed to*, which arose in the Early Modern English period and started spreading rapidly during the 18th century;
b. the complex conditional subordinator *supposing that*, which — like *seeing (that)* — seems to have been an ever-available but rarely chosen option for the past 400 years.

In the terminology proposed above, the modal idiom thus represents the dynamic type of grammaticalisation, whereas the complex subordinator illustrates the static type.

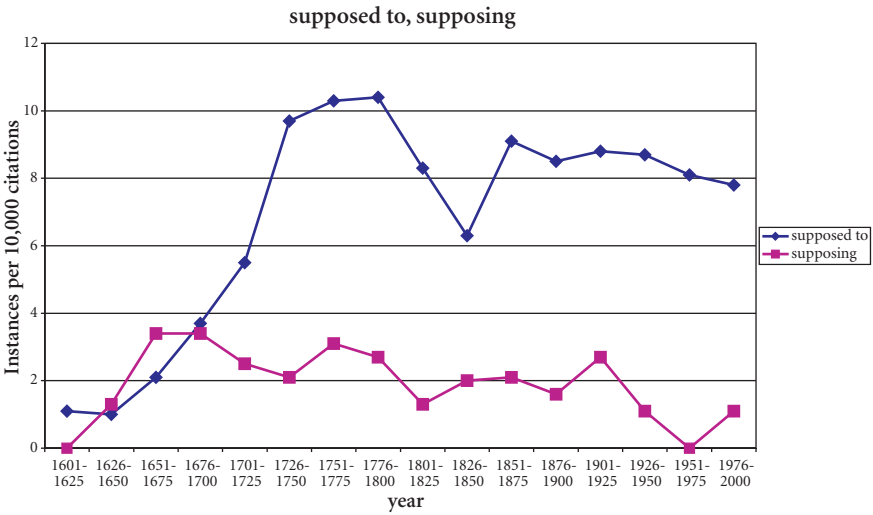


Figure 5. *Supposed to/supposing* 1600–2000 — frequency as *n*/10,000 citations

Figure 5 shows the relevant frequencies. The graph for *supposed to* shows precisely the type of delayed increase in frequency noted for *going to*, *begin/start*, and *help* above, and to this extent the example my serve as another textbook instance of "dynamic" grammaticalisation leaving a strong statistical imprint in a historical corpus. Even here, however, it pays to take an additional detailed look at the earliest (pre-1600) attestations of the construction. As is to be expected, the lexical use of *be supposed to*, as the passive equivalent of "somebody supposes something to be the case", is more strongly in evidence than it is today. While in the grammaticalisation of modal auxiliaries deontic meanings generally precede epistemic ones, the reverse seems to hold in the present case — clearly an effect of the epistemic meaning of the verb *suppose*, the lexical source.

Of 21 relevant pre-1600 examples, a mere three allow a deontic interpretation, and in only one instance is such an interpretation really the most likely one:

(11) To inquire⟨dd⟩what number of Acres, the place of Common, wherein the surcharge is supposed to be made, doth containe (1598 J. Manwood Lawes Forest xiv. 84)

On the evidence of the OED data, the complex conjunction *supposing (that)* exemplifies static grammaticalisation. There is no diachronic statistical trend, and the occasional grammatical use of the participle to introduce conditional clauses has been a stable but rarely chosen option for the past 400 years.

9. Conclusions

Focussing on several instances of variable usage in present-day English and tracing their history since ca. 1600, the present paper has elaborated on, and partly corrected, existing descriptive accounts. In addition, the results of the present investigation suggest a few methodological and theoretical generalisations which are worth formulating by way of a conclusion.

To corpus linguists interested in the history of the English language, the present paper has proved the value of the OED's quotation base as a historical corpus. The quotation base is particularly rich for the two centuries from 1800 to the present, which receive only unsystematic and partial treatment in "classical" historical grammars such as Jespersen (1909–1949) or Visser (1970–78). Even in recent works devoted specifically to the history of English in the nineteenth century, grammatical change does not figure as a prominent subject (Bailey 1996), or treatment is restricted to giving a conspectus of

existing research (Görlach 1999). Denison (1998) remains as the only survey of the recent history of English grammar which combines a review of the state of the art and original research. It is impressive in the breadth and depth of its coverage, but as a pioneering first effort it inevitably raises as many questions as it answers. Clearly the systematic exploration of the available corpora, and the OED quotation base in particular, for what they can tell us about recent and ongoing grammatical changes in English has only just begun, and much promising work lies ahead.

For the student of grammaticalisation, the most important general lesson to be learnt from the investigations reported on here is that ongoing grammaticalisation processes can be recognised and documented for languages such as English, for which there is a rich panchronic corpus linguistic working environment. It is understandable for programmatic and theoretical work on grammaticalisation to stick with well-worn study examples such as the *going-to* future and to illustrate the analyses with second-hand data from standard philological works of reference. But what grammaticalisation theory needs in order to progress is "liberation from secondary data" — as Fred Newmeyer (1998: 328) put it in a recent critique, which — coming from a generativist — was surprisingly sympathetic otherwise. I would argue that such liberation is brought about most easily through the use of corpora, which are becoming available in increasing numbers even for languages outside the Standard Average European compass.

Unsurprisingly, the analysis of corpora is particularly revealing with regard to the role of statistics in grammaticalisation. For the "dynamic" instances of grammaticalisation (that is the diachronically directed ones which leave a statistical imprint in discourse) the present study has shown that grammaticalisation is not accompanied by a simultaneous across-the-board increase in discourse frequency. Such increases, when they are in evidence, should rather be seen as a delayed symptom of earlier grammaticalisation having occurred and allow us to trace the spread of a new grammatical form through different styles and genres. Changes in relative or proportional frequencies, however, will usually be part of the central phase of the process of grammaticalisation itself. For example, as we saw in the analysis of *going to*, infinitives started crowding out prepositional complements rapidly right during grammaticalisation, and long before the overall frequency of *going to* took off on its dramatic rise. However, corpora provide insights into grammaticalisation processes that go beyond the statistical. Thus, through qualitative analysis of individual attestations the corpus has helped to document instances of "static" grammaticalisation which are independent of any statistical basis in discourse. This serves as a proof for

the continuing relevance of textlinguistic and qualitative analysis of corpus evidence[8] even in an age of ever-larger databases and ever more sophisticated statistics. In addition, it should make us think again about the limited role of statistics even in those cases in which it is involved.

Once we give up the bird's-eye view conferred by several centuries of hindsight and turn our attention to what went on in the minds of the speakers and listeners who actually brought about grammaticalisation, an inevitable question arises: what of the statistical shifts we observe in our corpora was salient in the minds of the speakers and listeners at the time? For what Suzanne Romaine has said about the role of statistics in stylistics is certainly no less true for grammaticalisation:

> How much of a difference is significant in terms of *perception* by the listener? One needs to know this to determine if the variable is being used symbolically in a successful way. Tests of statistical significance only indicate *production*-related differences. (personal communication, quoted in Rickford 1999: 150)

After all, grammaticalisation did not proceed because a form was becoming more and more frequent, but because successive generations of speakers perceived the phenomenon and developed it further (or kept an option for grammaticalisation alive in the "static" cases).

Statistics helps us chart the spread of known cases of grammaticalisation; given favourable circumstances, it may even help us identify ongoing processes of grammaticalisation that have not caught the attention of grammarians for some reason or other; but if we are to understand the causes of grammaticalisation, we first and foremost need a theory of communication grounded in general linguistics and cognitive psychology. And in order to contribute towards the building of such a theory, we need corpora for two reasons: sometimes because they provide a lot of data (which can be analysed statistically), but on other occasions because they provide authentic data (which allow us to analyse language in performance qualitatively without reducing its complexity).

While my bias as a corpus linguist has led me to try to "sell" corpora to students of grammaticalisation in the main body of the present paper, I would thus like to redress the balance in favour of linguistic theory at least in the conclusion. The advantage the present corpus-linguist has derived from grammaticalisation theory is easily stated but fundamental. Without it, the above remarks would be a "Zahlenfriedhof", a cemetery of numbers, — an incoherent compilation of uninterpreted and hence pointless statistics.

Notes

* The present paper has profited from comments and suggestions made by the participants at the CORGIE (Corpora and Grammaticalization in English) symposium in Växjö in April 2001, in particular Marianne Hundt, Hans Lindquist, Terttu Nevalainen, Sali Tagliamonte, and Gunnel Tottie. Bernd Kortmann (Freiburg) and Merja Kytö (Uppsala) generously commented on a previous written version.

1. Added consolation on that score is provided by Willinsky 1994, who evidently started out to perform a Foucault-inspired de(con)struction job on the OED but ended up as a grudging admirer. There is, it seems, relatively little systematic exclusion of textual witnesses, and what little there is — for example, the neglect of the politically radical literature emanating from the 19th-century English workers' movement — would not have made a difference to the language-historical picture (p. 188).

2. In the course of this project, Brown and LOB, the two standard one-million word corpora of written American and British English of 1961, were complemented with updates matching the originals as closely as possible in size and composition, but containing texts of 1992 and 1991 respectively. The Freiburg update of Brown is generally referred to as "Frown", the update of LOB as "F-LOB". For an overview of this project see, for example, Mair 1998 and Hundt and Mair 1999.

3. Here understood in a wide sense, as any grammatical process affecting the surface realisation of participant structure.

4. The oft-quoted "thys onhappy sowle .. was goyng to be brought into helle for the synne an onleful lustys of her body" of the Monk of Evesham. The example has the advantage of being a truly diagnostic one, as both the semantics of the verb *bring* and the passive of the infinitive are incompatible with a literal interpretation of the verb *go* as a motion verb. Note also that the *going to*-form is used in the past tense, illustrating the future-in-the-past use that has remained common to the present day. The example is a free translation from a Latin original, which, however, does not seem to have any direct impact on the use of *going to* (see Danchev and Kytö 1994:61 for a discussion). Danchev and Kytö (ibid.) cite a potential earlier case from 1438, in which the verb *go* could with some justification be read in its literal, motion-verb sense. On possible French influence on the rise of the construction in English, an interesting issue not directly relevant to the concerns of the present paper, see Danchev and Kytö 2001.

5. Part of the precise shape of the curve is undoubtedly due to the nature of the corpus, with its bias towards the written language. Assuming that the *going to*-future arose in the spoken language, and then started spreading into formal written discourse, we would expect a time lag, indicating the period it took for the new form to lose its "informal" stigma and to appear in writing.

6. "It is phenomenologically impossible to observe grammaticalisation actually unfolding in its earliest stages — that is in the formation of formulae and simple patterns. Vast masses of text containing potentially grammaticalising structures would have to be recorded and — what is more — analysed. But what would be the criteria? And which would be the structures?" (translation mine)

7. On this assumption, we would probably handle the instances assembled under (c) as examples of syntactic re-analysis without grammaticalisation. Note, though, that even König and Kortmann, who argue for re-analysis in the comparable case of English complex deverbal prepositions, conclude: "Grammaticalization, which to a certain extent the reanalysis of verbs as prepositions is, generally involves a loss of features: Constraints on word order variability, loss of agreement, phonological and morphoerosion as well as semantic bleaching" (1991:123).

8. Or, as some might put it, the close reading of old texts in the spirit of traditional philology.

Appendix: Statistical documentation for Figures 1 to 5

1. OED/number of quotations per quarter century: 1600–2000

Period	Number of quotations
1601–1625	118,285
1626–1650	85,783
1651–1675	100,018
1676–1700	79,212
1701–1725	73,341
1726–1750	56,520
1751–1775	64,884
1776–1800	78,998
1801–1825	111,374
1826–1850	174,073
1851–1875	222,261
1876–1900	256,498
1901–1925	104,297
1926–1950	122,313
1951–1975	208,900
1976–2000	46,182

2. OED/*going to*

	going to	*going to*/10,000 quotes	*gonna*	*gonna*/10,000 quotes
1601–1625	22	1.9		
1626–1650	31	3.6		
1651–1675	45	4.5		
1676–1700	39	4.9		
1701–1725	45	6.1		
1726–1750	44	7.8		
1751–1775	42	6.5		
1776–1800	62	7.9		
1801–1825	76	6.8		
1826–1850	121	7.0		
1851–1875	206	9.3		
1876–1900	306	11.9		
1901–1925	289	27.7	4	0.4
1926–1950	375	30.7	24	2.0
1951–1975	703	33.7	32	1.5
1976–2000	174	37.7	8	1.7

NB: Of the 1601–1625 exx., 6 are infinitives, some plausible *going to* futures.
Gonna is attested for 1913ff., with references to previous Scots *gaunna/ganna.*

3. OED/*start* to*[a]

	Total	Total/10,000 quotes
1601–1625	1[b]	0.1
1626–1650	3	0.3
1651–1675	3	0.3
1676–1700	1	0.1
1701–1725	2	0.3
1726–1750	2	0.4
1751–1775	1	0.2
1776–1800	4	0.5
1801–1825	1	0.1
1826–1850	7	0.4
1851–1875	13	0.6
1876–1900	23	0.9
1901–1925	26	2.5
1926–1950	52	4.3
1951–1975	121	5.8
1976–2000	42	9.1

[a] The asterisk represents the "any character" symbol ("wildcard") in the OED's search software. A search for *start** collects the forms *start*, *starts*, *started*, *starting* (and — unfortunately — also *startling*, *startled*, *starter*, and some others, which, however, are too rare to have a distorting influence). What is a problem, though, is the identity in form between the noun *start* and the verb, which means that frequent idiomatic expressions such as *from start to finish* will show up in the data. The results for the earlier periods are additionally distorted by the fact that *start* commonly governed a prepositional phrase with *to* (e.g. *start to school*). Distortions are severe for the 17th, 18th and 19th centuries. Thus, manual post-editing of part of the output reveals that for the third and fourth quarter of the nineteenth century 4 out of 7 and 10 out of 13 hits are spurious. Fortunately, the situation improves considerably for the 20th century, that is the period during which the statistical rise becomes apparent (spurious hits: 3 of 26 and 8 of 52 for the first and second quarters respectively).

[b] Obviously, this 1613 example is not an aspectual construction but instantiates the old meaning of *start* ('shrink back suddenly') with an adverbial infinitive:

a1613 Overbury *A Wife*, etc. (1614) H 4 b, The charitable man dreames of building Churches, but starts to thinke the vngodly Courtier will pull them down again.

Cf. similarly:

1664 Cotton *Scarron.* i. 47 And up he starts, to go a stealing, Either a Muttning, or a Vealing.

4. OED/*start** **ing*[a]

	Total	**Total/10,000 quotes**
1601–1625	3	0.3
1626–1650	–	–
1651–1675	–	–
1676–1700	–	–
1701–1725	–	–
1726–1750	–	–
1751–1775	1	0.2
1776–1800	5	0.6
1801–1825	–	–
1826–1850	4	0.2
1851–1875	4	0.2
1876–1900	6	0.2
1901–1925	27	2.6
1926–1950	81	6.6
1951–1975	212	10.1
1976–2000	63	13.6

[a] Again, most of the early instances are spurious, showing *start* followed by an adjective ending in *-ing*, or by a verbal participle that functions as an optional adverbial expansion of the clause rather than as a complement of the verb *start* in an aspectual construction.

5. OED/*begin*/began/begun to*

	Total	Total/10,000 quotes
1601–1625	265	22.4
1626–1650	218	25.4
1651–1675	270	27.0
1676–1700	230	29.0
1701–1725	225	30.7
1726–1750	166	29.4
1751–1775	230	35.4
1776–1800	216	27.3
1801–1825	292	26.2
1826–1850	465	26.7
1851–1875	491	22.1
1876–1900	671	26.2
1901–1925	307	29.4
1926–1950	326	26.7
1951–1975	578	27.7
1976–2000	131	28.4

6. *help*
OED/*help* (20th century): absolute frequencies

Decade	All relevant uses	*to*			Zero		
		Total	+NP	−NP	Total	+NP	−NP
1901–10	24	20	4	16	4	4	0
1911–20	21	14	7	7	7	6	1
1921–30	25	17	7	10	8	2	6
1931–40	48	32	15	17	16	9	7
1941–50	52	26	9	17	26	7	19
1951–60	91	47	17	30	44	17	27
1961–70	136	71	19	52	65	34	31
1971–80	190	85	24	61	105	52	53
1981–90	25	8	3	5	17	7	10

OED/*help* (20th century): normalised frequencies (= instances per 10 000 quotes)

Decade	Number of quotations	All rel. uses	*to*			zero		
			total	+NP	−NP	total	+NP	−NP
1901–10	52,085	4.6	3.8	0.8	3.0	0.8	0.8	0
1911–20	30,785	6.8	4.5	4.25	4.25	2.3	2.0	0.3
1921–30	47,699	5.2	3.6	1.5	2.1	1.7	0.4	1.3
1931–40	52,070	9.2	6.1	2.9	3.2	3.1	1.7	1.4
1941–50	44,091	11.8	5.9	2.4	3.5	5.9	1.6	4.3
1951–60	69,458	13.1	6.8	2.5	4.3	6.3	2.4	3.9
1961–70	90,015	15.1	7.9	2.1	5.8	7.2	3.8	3.4
1971–80	86,354	22.0	9.8	2.8	7.0	12.2	6.0	6.2
1981–90	9,410	26.6	8.5	3.2	5.3	18.1	7.4	10.7

OED/*help* (19th century): absolute frequencies

Decade	All relevant uses	*to*			zero		
		Total	+NP	−NP	Total	+NP	−NP
1801–25	20	18	5	13	2	2	0
1826–50	32	28	9	19	4	3	1
1851–75	44	40	18	22	4	1	3
1876–1900	85	75	20	55	10	7	3

OED/*help* (19th century): normalised frequencies (= instances per 10 000 quotes)

Decade	Number of quotations	All rel. uses	*to*			zero		
			Total	+NP	−NP	Total	+NP	−NP
1801–25	111,374	1.8	1.2			0.6		
1826–50	174,073	1.8	1.1			0.7		
1851–75	222,261	2.0	1.0			1.0		
1876–1900	256,498	3.3	2.1			1.2		

OED/*help* (18th century): absolute frequencies

Decade	All relevant uses	*to*			zero		
		Total	+Np	−NP	Total	+NP	−NP
1701–25	21	19	8	11	2	–	2
1726–50	14	13	5	8	1	–	1
1751–75	13	13	5	8	–	–	–
1776–1800	12	10	5	5	2	2	–

OED/*help* (18th century): normalised frequencies (= instances per 10 000 quotes)

Decade	Number of quotations	All rel. uses	*to*			zero		
			Total	+NP	−NP	Total	+NP	−NP
1701–25	73,341	2.9	2.6			0.3		
1726–50	56,520	2.5	2.3			0.2		
1751–75	64,884	2.0	2.0			0.0		
1776–1800	78,998	1.5	1.3			0.2		

OED/*help* (17th century): absolute frequencies

Decade	All relevant uses	*to*			zero		
		Total	+NP	−NP	Total	+NP	−NP
1601–25	34	27	12	15	7	3	4
1626–50	33	26	11	15	7	2	5
1651–75	25	24	7	17	1	–	1
1676–1700	21	19	7	12	2	1	1

OED/*help* (17th century): normalised frequencies (= instances per 10 000 quotes)

Decade	Number of quotations	All rel. uses	*to*			zero		
			Total	+NP	−NP	Total	+NP	−NP
1601–25	118,285	2.9	2.3			0.6		
1626–50	85,783	3.8	3.0			0.8		
1651–75	100,018	2.5	2.4			0.1		
1676–1700	79,212	2.7	2.4			0.3		

NB: 3 of the 7 "zero" examples from the second quarter (1626–1650) are from one text, namely Ward's *Simple Cobler*.

7.OED/*see* etc.
Absolute frequencies

	see	*sees*	*saw*	*seen*	*seeing*
1601–25	1152	44	236	112	118
1626–50	850	42	200	122	97
1651–75	857	44	304	287	132
1676–1700	659	36	275	225	57
1701–25	687	40	277	222	49
1726–50	464	57	153	178	45
1751–75	473	40	208	261	50
1776–1800	549	45	243	302	46
1801–25	863	54	458	505	78
1826–50	1140	57	547	685	93
1851–75	1482	106	692	895	129
1875–1900	1583	122	640	1000	144
1901–25	783	74	326	474	70
1926–50	902	54	351	554	86
1951–75	1331	103	554	877	134
1975–2000	355	29	118	205	36

Normalised frequencies/10,000 quotes

	see	*sees*	*saw*	*seen*	*seeing*
1601–25	97.4	3.7	20.0	9.5	10.0
1626–50	99.1	4.9	23.3	14.2	11.3
1651–75	85.7	4.4	30.4	28.7	13.2
1676–1700	83.2	4.5	34.7	28.4	7.7
1701–25	93.7	5.5	37.8	30.3	6.7
1726–50	82.1	10.1	27.1	31.5	8.0
1751–75	72.9	6.2	32.1	40.2	7.7
1776–1800	69.5	5.7	30.8	38.2	5.8
1801–25	77.5	4.8	41.1	45.3	7.0
1826–50	65.5	3.3	31.4	39.4	5.3
1851–75	66.7	4.8	31.1	40.3	5.8
1875–1900	61.7	4.8	25.0	39.0	5.6
1901–25	75.0	7.1	31.3	45.4	6.7
1926–50	80.2	4.4	28.7	45.3	7.0
1951–75	63.7	4.9	26.5	42.0	6.4
1975–2000	76.9	6.3	25.6	44.4	7.8

8. OED/*suppose*
Absolute frequencies

Period	Total	*supposed to*	*supposing*	*supposing that*
1601–1625	118,285	13	10	5
1626–1650	85,783	9	11	3
1651–1675	100,018	21	34	6
1676–1700	79,212	29	27	4
1701–1725	73,341	40	18	4
1726–1750	56,520	55	12	2
1751–1775	64,884	67	20	2
1776–1800	78,998	82	21	6
1801–1825	111,374	92	15	3
1826–1850	174,073	110	34	9
1851–1875	222,261	202	46	10
1876–1900	256,498	219	40	14
1901–1925	104,297	92	28	10
1926–1950	122,313	106	14	4
1951–1975	208,900	170	11	5
1976–2000	46,182	36	5	1

Normalised frequencies per 10,000 citations

Period	*supposed to*	*supposing*	*supposing that*
1601–1625	1.1	*	*
1626–1650	1.0	1.3	*
1651–1675	2.1	3.4	*
1676–1700	3.7	3.4	*
1701–1725	5.5	2.5	*
1726–1750	9.7	2.1	*
1751–1775	10.3	3.1	*
1776–1800	10.4	2.7	*
1801–1825	8.3	1.3	*
1826–1850	6.3	2.0	*
1851–1875	9.1	2.1	*
1876–1900	8.5	1.6	*
1901–1925	8.8	2.7	1.0
1926–1950	8.7	1.1	*
1951–1975	8.1	*	*
1976–2000	7.8	1.1	*

* = < 1.0

References

Bailey, R. W. 1996. *Nineteenth Century English*. Ann Arbor: University of Michigan Press.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brinton, L. J. 1988. *The Development of English Aspectual Systems: Aspectualizers and Post-Verbal Particles*. Cambridge: CUP.

Bybee, J. and Hopper, P. (eds.). 2001. *Frequency and the Emergence of Linguistic Structure*. Amersterdam: Benjamins.

Compes, I., Kutscher, S. and Rudorf, C. 1993. *Pfade der Grammatikalisierung: Ein systematischer Überblick*. Arbeitspapier 17. Köln: Sprachwissenschaftliches Institut der Universität.

Danchev, A. and Kytö, M. 1994. "The construction *be going to + infinitive* in Early Modern English". In *Studies in Early Modern English*, D. Kastovsky (ed.), 59–77. Berlin: Mouton de Gruyter.

Danchev, A. and Kytö, M. 2002. "The *go*-futures in English and French viewed as an areal feature". *NOWELE: Studies in North-Western European Language Evolution* 40: 29–60.

Denison, D. 1998. "Syntax." In *The Cambridge History of the English Language*, vol. IV: *1776–1997*, S. Romaine (ed.), 92–329. Cambridge: Cambridge University Press.

Givón, T. and Yang, L. 1993. "The rise of the English *get*-passive". In *Voice: Form and Function*, B. Fox and P. J. Hopper (eds.), 119–149. Amsterdam: Benjamins.

Görlach, M. 1999. *English in Nineteenth-Century England*. Cambridge: Cambridge University Press.

Hopper, P. J., and Traugott, E. C. 1993. *Grammaticalization*. Cambridge: CUP.

Huddleston, R. and G. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Hundt, M. 2001. "What corpora tell us about the grammaticalisation of voice in *get*-constructions". *Studies in Language* 25: 49–88.

Hundt, M. and Mair, C. 1999. "'Agile' and 'uptight' genres: The corpus-based approach to language change in progress". *International Journal of Corpus Linguistics* 4: 221–242.

König, E. and Kortmann, B. 1991. "On the reanalysis of verbs as prepositions". In *Approaches to Prepositions*, G. Rauh (ed.), 109–125. Tübingen: Narr.

Jespersen, O. 1909–49. *A Modern English Grammar on Historical Principles*. 7 vols. Copenhagen: Munksgaard.

Krug, M. 2000. *Emerging English Modals: A Corpus-Based Study of Grammaticalization*. Berlin: Mouton de Gruyter.

Lehmann, C. 1991. "Grammaticalization and related changes in contemporary German". In *Approaches to Grammaticalization*, Vol. II, E. C. Traugott and B. Heine (eds.), 493–535. Amsterdam: Benjamins.

Lightfoot, D. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.

Lightfoot, D. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Malden, MA: Blackwell.

Mair, C. 1994. "Is *see* becoming a conjunction? The study of grammaticalisation as a meeting ground for corpus linguistics and grammaticalisation theory". In *Creating and Using*

English Language Corpora, U. Fries, G. Tottie and P. Schneider (eds.), 127–137. Amsterdam: Rodopi.

Mair, C. 1995. "Changing patterns of complementation, and concomitant grammaticalisation, of the verb *help* in present-day British English". In *The Verb in Contemporary English: Theory and Description*, B. Aarts and C.F. Meyer (eds.), 258–272. Cambridge: Cambridge University Press.

Mair, C. 1997. "The spread of the *going-to*-future in written English: A corpus-based investigation of language change in progress". In *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak on his 60th Birthday*, Vol. II, R. Hickey and S. Puppel (eds.), 1537–1542. Berlin: Mouton de Gruyter.

Mair, C. 1998. "Corpora and the study of the major varieties of English: Issues and results". In *The Major Varieties of English: Papers from MAVEN 97, Växjö 20–22 November 1997*, H. Lindquist, S. Klintborg, M. Levin, and M. Estling (eds.), 139–157. Växjö: Acta Wexionensia.

Newmeyer, F.J. 1998. *Language Form and Language Function*. Cambridge, MA: MIT Press.

Olofsson, A. 1990. "A participle caught in the act: On the prepositional use of *following*". *Studia Neophilologica* 62: 23–35, 129–149.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rickford, J.R. 1999. *African American Vernacular English: Features, Evolution, Educational Implications*. Oxford: Blackwell.

Rissanen, M., Kytö, M. and Heikkonen, K. (eds.). 1997. *Grammaticalization at Work: Studies of Long-Term Developments in English*. Berlin: Mouton de Gruyter.

Thompson, S.A. and Mulac, A. 1991. "A quantitative perspective on the grammaticalization of epistemic parentheticals in English". In *Approaches to Grammaticalization*. Vol. II, E.C. Traugott and B. Heine (eds.), 313–330. Amsterdam: Benjamins.

Traugott, E.C. and Heine, B. (eds.). 1991. *Approaches to Grammaticalization*. 2 vols. Amsterdam: Benjamins.

Willinsky, J. 1994. *Empire of Words: The Reign of the OED*. Princeton, NJ: Princeton University Press.

Visser, F. Th. 1970–1978. *An Historical Syntax of the English Language*. 3 vols. Leiden: Brill.

Grammaticalisation from side to side

On the development of *beside(s)*

Matti Rissanen
University of Helsinki

The purpose of my paper is to describe the development and grammaticalisation of the preposition and conjunct *beside(s)* in English. This linking word probably goes back to the uncommon Old English construction *be + sidan*. In Middle English, the grammaticalisation of the preposition and conjunct takes place rapidly. This process involves the development of more abstract senses. Subjectification probably played a role in this development.

Two Old and Early Middle English syntactic constructions, i.e., dat. + prep. phrase (*him on hype* 'him on hip', 'by his side') and *þær* 'there' + adverb/preposition, (*þærin(ne)* 'therein'; *þærfore* 'therefore'), probably contributed to the development. The discussion is illustrated by evidence derived from available computerized corpora and historical dictionaries.

1. Introduction

The development of the English syntactic linking elements or connectives, i.e., prepositions, conjunctions and conjuncts,[1] illustrates the process of grammaticalisation from many angles. The nouns and verbs lying behind the grammaticalised items can be native or borrowed; the starting-point can be either a single lexical item or a group of words; the period in which grammaticalisation takes place may vary from Old to Modern English; and the parallel or differing sense developments of the original lexical items and the grammaticalised forms provide material for more general semantic considerations.

The purpose of this paper is to describe the grammaticalisation and semantic development of the preposition, adverb and (now obsolete) subordinator *beside(s)*.[2] The Middle English expansion of the senses of the noun *side* and the grammaticalised *beside(s)* will be outlined, with reference to the development of

the meaning of *side* from 'part of human or animal body' and other concrete locative uses to senses indicating abstract relations. Attention will be paid to the role of subjectification in the process of grammaticalisation: how the changing point of view of the speaker may have triggered the sense development of *beside(s)* from immediate vicinity to addition and even separation.

Paul Hopper's and Elizabeth Traugott's studies[3] offer an excellent basis for all studies of grammaticalisation, and their often-quoted definition of this process as one "whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions, and, once grammaticalised, continue to develop new grammatical functions" (Hopper and Traugott 1993: xv) efficiently defines this many-faceted phenomenon in diachronic linguistics. Bernd Heine's and Tania Kuteva's comparative *World Lexicon of Grammaticalisation* (2002) offers a comparative/cognitive perspective to the grammaticalised uses of the basic source concept 'side' in various languages and cultures. Heine and Kuteva's description (2002: 272–273) is, however, restricted to the concrete extensions of the senses of 'side': (1) beside; (2) locative; (3) near. The authors do not include the rich field of more abstract metaphorical and grammaticalised uses of the concept in their survey.

One purpose of this paper is to show how computerised corpora and databases can illustrate the long-term semantic development and various stages of grammaticalisation of a lexical item in English. The *Helsinki Corpus of English Texts* (HC) and the *ARCHER Corpus* give an overview of the development from the beginnings of English to the 1990s. This picture can be supplemented by details from other historical corpora, such as the *Dictionary of Old English Corpus* (DOEC), the *Middle English Compendium* (MEC), containing the *Middle English Dictionary* database (MED) and the *Corpus of Middle English Prose and Verse* (CME), the *Corpus of Early English Correspondence Sampler* (CEECS), and the *Century of Prose Corpus* (COPC). Present-day English evidence is derived from the *Lancaster-Oslo/Bergen Corpus* (LOB) and the *Brown Corpus*, from their Freiburg counter-parts of the 1990s (FLOB and Frown), and from the massive *British National Corpus* (BNC).[4]

The earliest evidence of the grammaticalisation of *beside(s)* goes back to Early Middle English and the process seems to be completed in Early Modern English, in the sixteenth or seventeenth century, although the final refinement in the syntactic use and sphere of senses between the shorter form *beside* and the longer form *besides* takes place in Late Modern English. Thus, this connective falls within the general pattern of the Middle English enrichment of the English grammatical system of connectives that can serve both as prepositions,

adverbs and subordinators. This enrichment may have been connected with the gradual typological change of English towards analyticity. It seems, too, that the need for grammatical means for expressing abstract relations between concepts and propositions increased in Middle English with the explosive growth in the number of vernacular texts representing new genres.

The long development of *beside(s)* has resulted in the following prepositional and adverbial uses of the word in present-day English (examples from FLOB and Frown):

(1) Called in again at St Joseph's on my way home, and was pleasantly surprised to find Mrs Knoepflmacher sitting **beside** Daddy's bed, in a bright yellow muu-muu and gold sandals. (FLOB K07:21)

(2) it would divert internal social and, as it was believed, pre-revolutionary pressures caused by the rapid industrialisation of the country, thus enabling them to preserve their power and status **beside** the Kaiser at the head of the empire. (FLOB G30:6)

(3) But whether or not it is true that it wouldn't matter to us is **beside** the point. (FLOB G63:14)

(4) My mother was **beside** herself with curiosity. (Brown R07:158)

(5) Even the memory of the punch that had sent Matthew to the ground was unimportant **beside** the fact that John was safe. (FLOB P03:16)

(6) Hence it is not surprising that no reaction products **beside** fragmentations are observed in the 1:2 clusters. (Frown J06:22)

(7) **Besides** Sting, artists he has worked with include Tanita Tikaram, Black and Lloyd Cole: languorous ballads and moody AOR for the CD generation. (FLOB E35:33)

(8) He was just over 65 and he reckoned she could not be much less than 60. Also she was intelligent and amiable **besides** being good-looking. (FLOB L23:14)

(9) I bet the first things to enter your mind at the mention of Spanish wine are rough, cheap plonk and Rioja — in fact, the two extremes of the business. But there is a great deal more **besides**, as we are beginning to discover. (FLOB E20:26)

(10) Although his English was good, there had always been few opportunities to practise of late; **besides**, he must remember that he had finished his education at Landsberg of all places, not at Oxford as had originally been intended. (FLOB K24:5)

In Present-day English, *beside* is used only as a preposition governing a noun or pronoun, both in concrete and more abstract senses (examples (1)–(6)). Example (3) and (4) are of interest for the present discussion, as the idiom indicates distancing instead of closeness. In example (5) the preposition indicates comparison, with little or no idea of local positioning left, and in (6) *beside* enters into the domain of *besides*, with the sense 'in addition'. This sense is obvious in examples (7) to (10), in which *besides* is used as a preposition governing a noun (7) and a non-finite clause (8), and as an adverb (9–10).

Even this short collection of examples indicates that the two forms have a fairly clear distribution of senses: the concrete indication of locality is expressed by *beside* and the more abstract relations mainly by *besides*. As will appear from this paper, in earlier periods of the language the semantic distribution of the two forms was much less clear-cut and both forms could be used prepositionally and adverbially and even to govern a subordinate clause.

2. Old English background

Side is most obviously one of the basic source concepts, comparable to *head*, *back*, *foot*, etc. (cf. Heine, Claudi and Hünnemeyer 1991: 32–33; 123–125). It is not surprising that by far the most common sense of the noun *side* in extant Old English texts is 'side of human or animal body' (cf. OED, s.v. *side* sb.[1] I.1; Bosworth-Toller, s.v. *side* an. I). Of the more than 250 examples appearing in the *Dictionary of Old English Corpus*, no less than 208 have this meaning.[5]

(11) wiþ **sidan** sare betonican leaf geseoð on ele & gebryte, alege on þa sidan (DOEC Bald's Leechbook 396)
'against the pain in the side, leaves of betony, boil in oil and break up, put on the side'

(12) ure Drihten sylf eowað us sona his blodigan **sidan** & his þyrlan handa & ða sylfan rode þe he for ure neode on ahangen wæs, (HC Wulfstan's Homilies, HOM WULF2 121)
'our Lord himself will soon show us his bloody side and pierced hands and the very cross on which he was hung for our need'

When *side* refers to an inanimate object, the referent is in most cases a building, boat (Noah's Ark!), or mountain. In some passages, *side* indicates the immediate neighbourhood rather than a part of the object; this can be regarded as the

first step towards the generalisation of the meaning of the word and as the beginning of its grammaticalisation when connected with various prepositions:

(13) On ane healfe þæs mynstres wæs an ormæte clif ascoren rihte adune, and seo deope ea Liger gehaten læg **on oðre sidan**, swa þæt man ne mihte to þam mynstre cuman butan þurh ænne pæð (DOEC Ælfric's Lives of Saints, St. Martin 77)
'On one side of the monastery there was a huge cliff descending right down, and the deep river called the Liger was on the other side, so that nobody could come to the monastery except through a path'

Note the correlation *on ane healfe … on oðre sidan* in this example.

There are no Old English instances of the use of *side* in clearly abstract contexts. These senses seem to develop in Middle English, alongside with the grammaticalisation of *beside(s)*.

The *Oxford English Dictionary* derives *beside* from the Old English phrase *be sidan* (prep. + dat. sg. of *side*). The etymology is plausible although this combination is practically non-existent in extant Old English texts. A simple comparison of the Old English Bible translations with the readings of the *King James Bible* shows that Old English had other means of expression for both the concrete and abstract senses of *beside(s)*:

(14) ða sende Iosue sceaweras to Hai, ðe ðær **gehende** wæs, (HC Ælfric, Joshua ALFOLD 7.2)
'Then sent Joshua spies to Ai, that there near was,'
"And Joshua sent men from Jericho to Ai, which is **beside** Beth-aven" (King James Bible)

(15) & þær wæron ofslagene feowertyne ðusenda of ðæs folces mæniu & seofonhund manna **butan** ðam ðe ðær ofslagene wæron on Chores ceaste. (HC Ælfric, Numbers ALFOLD 16.49)
'and there were killed fourteen thousand of the people and seven hundred except those that there were killed in Korah's strife.'
"Now they that died in the plague were fourteen thousand and seven hundred, **beside** them that died about the matter of Korah." (King James Bible)

The *Dictionary of Old English Corpus* gives only one example in which *be* is followed by *sidan/siden* without any intervening elements:[6]

(16) Sume hie ne mihton mode oncnawan
þæt wæs se deora Didimus wæs haten
ær he mid hondum hælend genom
sylfne **be sidan** þær he his swat forlet;
(DOEC Christ and Satan 543)
'[…] before he had touched Christ by the side where he shed his blood'

The value of this example is of course diminished by its occurrence in verse.

The combination of Old English *side*, without intervening elements, is extremely rare even with other prepositions; there are a few examples of *on/þurh/æt/fram* + *sidan*, but only in verse or in interlinear glosses:

(17) Oft mec isern [iron] scod sare **on sidan**; (DOEC Riddle 72,3)

(18) Wiht [creature/thing] is wrætlic [wondrous] þam þe hyre wisan [ways] ne conn.
Singeð **þurh sidan**. (DOEC Riddle 70,1)

(19) Vestiti dormiant et cincti cingulis aut funibus et cultellos suos ad latus non habeant dum dormiunt […] gescridde hi slapan & begyrde gyrdelsum oððe strængum & seax heora **æt sidan** & hi nabban þonne hi slapað þurh swefn (DOEC Benedictine Rule 264)

(20) O clerice, ne dempseris umquam dipticas lateri Eala ðu cleric ne wana ðu æfre wexbreda **fram sidan**. (DOEC Abbo of St. Germain, Bella Parisiacae Urbis III, London, 1)

In Old English the construction indicating 'by the side of X' was formed with *be* + gen./poss.pron. + X, as in the following examples:

(21) & befoh utan mid golde: & sting ut ðurh ða hringas **bi ðære earce** [**ark's**] **sidan**, ðæt hie man mæge beran on ðam, (DOEC Gregory the Great, Pastoral Care 169)

(22) Hy gewædode and begyrde resten and nane sex [sword] **be heora sidan** næbben, þelæs þe hy on slæpe gewundade weorþan, (DOEC Benedictine Rule 343; cf. (19), above)

Cf. the Present-day English, "I sat down by her side" (*Collins Cobuild English Language Dictionary*, s.v. **side** 1).

Although the seeds of *beside(s)*, with the necessary generalisation of the meaning of *side*, were probably sown in Old English, it is obvious that the story of the grammaticalisation of the preposition and adverb began in the Middle English period.

3. Middle English development

In the course of the Middle English period the sphere of meaning of the noun *side* shifts from 'part of human body' to locality in relation to inanimate objects. As can be seen in Table 1, in the first Middle English sub-period (1150–1250) of the *Helsinki Corpus* instances with reference to human or animal 'side' still prevail, while in the later sub-periods the figures are roughly equal, or the non-human reference prevails (ME3).[7]

Table 1. The noun *side* with animate and inanimate referents in the Middle English part of the Helsinki Corpus (absolute figures)

		Animate	Inanimate
ME1	(1150–1250)	8	2
ME2	(1250–1350)	17	13
ME3	(1350–1420)	19	71
ME4	(1420–1500)	43	38

In the course of the Middle English period the senses of *side* are extended from concrete to abstract. A selection of the definitions of the senses in the *Middle English Dictionary* (s.v. *side* n.), with reference to the dates of the earliest occurrences, will illustrate this development.[8]

1. A side of the human body, or of a part of the body, esp. between the shoulder and the thigh (2a. 1130–35)
2. A side of something; lateral face; the wall of a city or a room (1a. c1230)
3. Side by side (2f. c1275)
4. On every side (7a[b]. c1275)
5. In topographical or geographical terms (3b. c1300)
6. An area on one or another side of a topographical feature (6a. a1300)
7. Interest, concern; regard, point of view (9. c1275)
8. One of two parties to a dispute, contract, negotiation, contest, etc.; one of two contrasting groups (8b. a1300)
9. Fig. One of two contrasting or complementary abstractions, positions, dispositions, eventualities, etc. (5b[a]. 1340)
10. To the side, aside, away (7a[c]. c1400)

The line of development from part of the body (1,3), through concrete inanimate objects (2–6) to more abstract senses (7–10) can be clearly seen. Below,

MED examples of the abstract uses are given (23)–(27).

Sense 7, 'point of view'

(23) Al so þu dost on þire **side**. (c1275 MED Owl & N., Clg A.9, 429)

Sense 8, 'party in dispute'

(24) [N]ou crist ous lete so bitide [Þ]at we ne wallen in þar **side**. (?a1300 MED Fiftene toknen, Dgb 86, 136)

Sense 9, 'position', 'aspect'

(25) Huanne þise tuo **ziden** of þe herte byeþ acorded and y-ordayned, þet is þe scele and þet wyl, (1340 MED Ayenb., Arun 57, 153/14)

(26) He mai lihtly be deceived That tristith unto mannes helpe; Bot wel is him whom god wol helpe For he stant on the siker **side**. (a1393 MED Gower CA, Frf 3, 5.2427)

Sense 10, 'away'

(27) Feith had first siȝte of hym, ac he flegh on **syde**. (c1400 MED PPl.B, LdMisc 581, 17.57)

The grammaticalisation of *beside(s)* follows the same line from human/animate through concrete inanimate to abstract reference.[9] The earliest example (28) is recorded from the mid-twelfth century in the *Middle English Dictionary* database (s.v. *paradise* 1a).[10]

(28) Seth wuneda on ana munte **beside** paradise (c1150 MEC MED Annot. Cld. OT, Cld B.4)

There is ample evidence of the grammaticalised forms in thirteenth century texts, both as a preposition (29)–(31), as an adverb (32)–(33), and in post-position (34):

(29) On þe niht […] were herdes wakiende **bi side** þe buregh. (a1225 MED Trinity Homilies 14. 52)

(30) Þenc as tah he heng **biside** þe blodi up o rode. (a1250 MED Wooing Lord, Tit D.18, 287)

(31) þat he lette makien ænne dich; þe wæs wunderliche deop.
muche and swiðe stronge; **bisides** Scotlonde.
from sæ to sæ eode [went] þæ dich
(c1275 MED Layamon's Brut, Clg A.9, 5162)

(32) Ȝif æuere æi [any] of þine gume [men]; gære [spear] haueð **bisiden**.
send hit mid freond-scipe; feor from us seoluen.
(c1275 MED Layamon's Brut, Clg A.9, 7596)

(33) Æfne þan worde; þe þe cniht sæide. of Oxene-uorde.
to his iueren **bisiden**; þe gon he to riden;
æfne al swa swiðe; swa hund þene heort driueð.
(c1275 CME Layamon's Brut, Clg A.9, 13356)

(34) Sune min so leue, site nu me **bisiden**.
(a1250 MED Proverbs of Alfred, Glb. A.19, 129/539)

The *Middle English Dictionary* database records 19 thirteenth-century instances of this type in seven texts. The syntactic constructions displayed by the above examples show that the grammaticalisation of *beside(s)* was well under way as early as the 13th century. Forms with or without the final *-n* or *-s* can be found, and there is variation even in the same text, as can be seen from examples (31)–(33). All the instances quoted above show concrete spatial relationship, with reference to the human body or to a place.

The *Middle English Dictionary* gives an abstract interpretation, 'in addition', 'besides', 'moreover', to three passages from Layamon's *Brut*:

(35) Heo letten forð **bi-siden**; an oþer folc riden.
ten þusend kempen; (c1275 CME Lay Brut, Clg A9, 2742)

(36) and al þat folc Frensce; bihongen weoren feire.
i-wepned wel alle; and hors ho hafden uatte [fat].
Þer weoren **bisides**; fiftene biscopes.
(c1275 CME Lay Brut, Clg A9, 12182)

(37) & ælc bær an honde; anne saȝel stronge.
and **bisiden** heo gunnen heongen; cniues swiðe longe.
(c1275 CME Lay Brut, Clg A9, 6128)

Example (37) is misinterpreted by the MED: the passage means 'they hung long knives by their sides', cf. (32) above. Even in (35) and (36) the meaning is probably concrete, 'by the side (flank) of other groups of people', but the route towards more abstract senses of the type 'in addition' is traceable from examples of this kind.

The adverbial constructions formed with *there* + *beside(s)* appear in the thirteenth century, too:

(38) A sunne ful forȝeuelich mei wurðe ful deadlich þurh sum uuel totagge
[circumstance] þe lið **þer bi siden**. (c1230 MED Ancr., Corp-C 402, 177/9)

(39) Þo stod on old stoc **þarbiside** (c1275 MED Owl & Nightingale, Clg A.9, 28)

Here + *beside(s)* occurs from c. 1300 on:

(40) A kniȝt þer was in Engelond, by norþe **her biside.** (c1300 MED South English Legendary, Oxford Scholar, Hrl, 1)

On the basis of available textual evidence it is impossible to determine the mutual order of the emergence of the prepositional and adverbial uses of *beside(s)*. It would seem natural, however, that the adverbial use of the compound developed first. This would also be in accordance with the general line of development of Old English prepositions. A possible first step from the Old English noun phrase *be* + gen./poss.pron + *sidan* might be the construction, not uncommon in Old English, in which possessive relationship is indicated with the dative form of the personal pronoun + a prepositional phrase. Mitchell (1985: §304, p. 124; cf. also §1357) gives the following example:

(41) and Abraham hafde **him on handa** fyr and swurd (Ælfric, Catholic Homilies II 544)

Cf. also the Old English example (16) (*hælend* […] *be sidan*) and the Middle English examples of *þe* […] *besiden* (fn 7). and *me* […] *bisiden* (34), which could be translated as 'by Christ's/thy/my side'.[11]

A construction possibly supporting the development of the preposition *beside(s)* from the adverb in Early Middle English is the combination of *there* + *be* + *siden*, see (38)–(39), above. This pattern could easily be formed on the model of the construction pers.pron.(dat.) + *be* + *sidan* described above. As Österman (1997) has shown, the compound adverbs formed with *there* and a preposition (*therefore, thereto, therewith*, etc.) increased rapidly in the Early Middle English period. A reanalysis of *there* + *beside(s)* to *beside(s)* + *that* would not be unnatural, following the pattern of *thereto/to that*, *therefore/for that*, etc.

The development from concrete to abstract senses of *beside(s)* is evidenced from the fourteenth century on. This development can be traced in the senses listed in the *Middle English Dictionary* for the adverb and preposition *biside(s)*:

Adverb:
1. At the side, at one's side; alongside
2. Nearby, in the neighborhood; nigh, faste ~; her ~, ther ~; passen ~, pass by
3. (a) To one side, aside; away; blenchen ~, to dodge; flen, gon ~, run away, escape; avoid; leien, putten, setten ~, put or set aside, dismiss; leven ~, leave

alone, shun; sheden ~, spill over the sides (of a spoon); (b) gon ~, go astray; ?miss an opportunity, go or do without
4. In addition, besides, moreover
5. (a) Of time: near; (b) of rank: on a level, equal.

Preposition:
1. (a) At or by the side of, next to; (b) alongside of, along; (c) in company with, along with; in the presence of
2. (a) Near or close to; faste ~; of biside (a place), from the vicinity of; (b) (a certain distance) from (a place or person)
3. To one side of; loken ~, look to (one's) side, look aside; gon ~, pass (sth.) by, forego; gon ~ the wei, stray from the path, go astray
4. In addition to, along with, over and above
5. (a) Outside (a place); outside the bounds of (true doctrine, truth, reason, nature); beyond the control of (one's will); also, contrary to; gon ~, deviate or stray from; (b) ~ leve, without permission or consent; (c) ~ himself, out of his wits

Senses 1 and 2 of the adverb and 1 and 2(a) of the preposition correspond to the basic human and concrete non-human references of the noun *side*. Senses 3(a) of the adverb and 2(b), 3 and 5(a) of the preposition have moved away from the senses of 'part of', or 'in the immediate vicinity to' and contain an implication of distancing, movement away, in either the concrete or abstract sense (cf. sense 10 of the noun *side*, ex. 27, above). Senses 3(b), 4 and 5 of the adverb, and senses 4 and 5 of the preposition represent abstract uses. The following are some MED examples of the distancing and abstract uses:

Adverb 3(a) 'aside', 'away'

(42) Arthur teh [turned] **bi-side**; and saide to his iveres. (c1300 MED Layamon's Brut, Otho C.13, 12982))

(43) Al it passeth thurgh myn Ere..And is foryete and leid **beside**. (a1393 MED Gower CA, Frf 3, 2.1993)

Preposition 3, 5(a) 'aside', 'outside'

(44) The glorie of hym is to go **biside** [L praetergredi] wicke thingys. (a1382 MED WBible1, Dc 369[1] Prov.19.11)

(45) One naturale, þe toþer **biside** nature; One materiale, þe toþer formale. (?a1425 MED Chauliac1, NY 12, 46b/a)

Adverb 3(b), 4, 5 'astray', 'moreover', 'near', 'equal'

(46) Som man..hath thing that mai him plese..Where as I faile and go **beside**. (a1393 MED Gower CA, Frf 3, 4.2862)

(47) Of þe lond of france, & of oþer lond **bi syde**. (c1325 MED Glo. Chron.A, Clg A.11, 2053)

(48) **Bisyde** is þe day of perdicioun. (c1450 MED Interpol.Rolle Ps., Bod 288, 44)

(49) Hi yelt..loue to ham þet byeþ **bezide**, grace to ham þet byeþ beneþe. (1340 MED Ayenb., Arun 57, 126)

Preposition 4, 5(b–c) 'in addition to', 'outside', 'beyond'

(50) Fier is on hem **bi-siden** ligt; fele it brende and made o-frigt. (a1325 MED Gen.& Ex., Corp-C 444, 3651)

(51) Thogh it be **beside** hire leve, I hope..That I do noght ayein the pes. (a1393 MED Gower CA, Frf 3, 3.530)

(52) And as a man **beside** hem self be farde. (a1500 MED Gener.2, Trin-C O.5.2, 4786)

It seems that the cognitive process that causes the extension of meaning from nearness to distancing can be related to subjectification and to the changing point of view. As long as the point of view is an outsider's or neutral, the idea of 'side' most naturally refers to somebody or something in the immediate vicinity of the person or object governed by *beside(s)*. But when the relation is defined from the point of view of this referent, distancing, movement away, becomes a natural extension of meaning. With this development the way is paved for the emergence of abstract meanings — not only 'in addition to' but also 'outside', 'except', etc.[12]

In the course of the Middle English period, the number of occurrences of *beside(s)* shows a considerable increase.

As can be seen in Table 2, the Early Middle English sub-periods ME1–2

Table 2. Occurrences of *beside(s)* in the Middle English part of the Helsinki Corpus

		n	**Per 10,000 words**
ME 1–2	(1150–1350)	15	0.7
ME 3	(1350–1420)	40	2.2
ME 4	(1420–1500)	26	1.2

(1150–1350) of the *Helsinki Corpus* give 15 instances of *beside(s)*, all from thirteenth or early fourteenth-century texts (Layamon's *Brut* 3, *King Horn* 4, Robert of Gloucester's *Chronicle* 4, *South-English Legendary* 1, *Historical Poems* 2, *Metrical Psalter* 1). All the instances refer either to the human 'side' or define closeness or immediate neighbourhood to a concrete object, most often a geographical location (town, river, way, etc.). In late fourteenth and early fifteenth-century texts (HC ME3, 1350–1420), the number of examples is almost trebled (40 instances). The prepositional use, in particular, gains ground rapidly: while in the ME1–2 texts *beside(s)* is used as a preposition in about half the instances, this is the case in no less than thirty-three of the 40 instances in ME3. In this sub-period, too, four occurrences of the abstract use of *beside(s)* can be found. The most interesting of these are the following, in which the adverb *beside* is linked with *aught* or *enough* and governed by the verb *have*. Here, obviously, the grammaticalisation of the phrase is completed (cf. the *Middle English Dictionary* entry, adverb, sense 3 and example (43), above).

(53) And a ryche man hyt noyeþ oftyn tyde
þat a pore man hat oght **besyde**.
Alle þat he may, with euyl he fondys
For to reue hym, and haue hys londys;
(c. 1400 HC Handlyng Synne HS 194)

(54) ȝif a man schulde wenden aweye, hit were no nede to chargen hym wiþ þingus þat weren not profiȝtable, ȝif he hadde ynow **bysyde**. (c1400 HC Wyclif, Sermons WSERM I 357)

Amongst the 26 *Helsinki Corpus* instances of *beside(s)* recorded from the later fifteenth century, abstract senses are proportionately more frequent and varied. The earliest instance of *beside one's wit*, which may lie behind the expression *beside oneself* (MED, prep., sense 5c, example (52), above), is worth quoting:

(55) And thenne they enteryde in to the cytte of London as men that hadde ben halfe **be-syde** hyr wytte; and in that furynys they wente, as they sayde, for the comyn wele of the realme (c1475 HC Chronicle of London CHRLOND 191)

The *Corpus of Middle English Prose and Verse* (CME), part of the *Middle English Compendium*, is much more extensive than the *Helsinki Corpus*. There are, all in all, almost six hundred occurrences of *beside(s)*. As this corpus, unlike the *Helsinki Corpus*, consists of complete texts, it gives interesting information on the frequency in individual texts.[13] To give an example, while all the 59 instances of

beside(s) refer to location in the narrative and fairly archaic text of Malory's *Le Morte Darthur*, in the *Paston Letters*, consisting of writings of several people, all instances have abstract reference indicating addition (ex. 56), or, in negative contexts, exception (57).

(56) All thes shwld haue holpyn me well þer-to, **by-syde** othyre thyngys that I haue boryn these yerys þat I speke not of. (1471 CME Margaret Paston to John Paston III 11, 05)

(57) that there is owt off that contré that be nat at Norwych **besyde** me, that be ryght worshypfull (1470 CME John Paston II to John Paston III 11, 15)

The very frequent use of *beside(s)* in the *Paston Letters* written in the fifteenth century (32 instances) might suggest that the expansion of the sphere of senses of this word towards more abstract domains, with a new prototypical centre of 'in addition', 'other than', was initiated at the informal spoken level of the language, as a change from below. This suggestion is supported by the infrequent use of *beside(s)* in the statutes and official documents (there are no instances in the *Helsinki Corpus* material), which are rich in other new connectives, particularly those borrowed from French or Latin (see, e.g. the studies quoted in fn. 2 above).[14]

4. Modern English establishment

In the Modern English period, from the sixteenth century to our days, the variety of abstract senses further increases. Of the senses listed in the *Oxford English Dictionary* (s.v. *beside* and *besides*) with first occurrences in sixteenth- and seventeenth-century texts, the following illustrate this development:

Beside (adv.) 3; *besides* (adv.) 3: Other than, else than, otherwise

(58) And one day in the week to touch no food, And but one meal on euery day **beside**. (1588 OED Shakes. L.L.L.I.i.40)

(59) Which..leaves behinde a stayne Upon the beautie of all parts **besides**. (1596 OED Shakes. 1 Hen. IV, III. i. 185)

Beside (prep.) 4c; *besides* (prep.) 4b,c: of removal, deprivation: out of, away from

(60) Least he should be set **beside** the kingdome which he..held. (1548 OED Udall, et., Erasm. Par., Matt. ii. 25)

Table 3. *Beside(s)* in the Early Modern English part of the Helsinki Corpus (EModE1 = 1500–1570; EModE2 = 1570–1640; EModE3 = 1640–1710)

	Preposition				**Adverb**			
	beside		*besides*		*beside*		*besides*	
	Local	Abstr.	Local	Abstr.	Local	Abstr.	Local	Abstr.
EModE1	6	17	4	18	1	3	0	4
EModE2	1	3	2	21	1	9	1	13
EModE3	1	3	0	24	0	0	0	24

(61) Thou mayest well, **besides** Christ, know him [God] as a tyrant. (1537 OED Tindale, Exp. 1 John, Wks. II. 183)

(62) The husbandmen..by coueyne and fraud..be put **besydes** it. (1551 OED Robinson, tr. More's Utop. 41)

All in all, in Modern English the various abstract uses of *beside(s)* become more common, we could even say more (proto)typical, than the concrete local senses. The rough distribution of the occurrences of the concrete and abstract uses in the Early Modern English (1500–1710) part of the *Helsinki Corpus* can be seen in Table 3.

The number of the occurrences of the local use of both forms is very low. It is also obvious that the longer form, with the morphological adverbial marker *-s*, gains ground at the cost of the shorter form, even in prepositional uses.[15]

In view of the general lines of development of adverbial connectives it is not surprising that in the sixteenth century *beside(s)* can be used as a subordinator, examples (63)–(65) and fn. 16. This use is, however, rare (four instances in the *Helsinki Corpus* material) and was never established in the language.[16] The one instance in which *beside* alone introduces a subordinate clause (65), is of particular interest:

(63) Sire **besids that** I am your Graces subject and servant, and sworn off your counsel thoff unworthi, your Grace hath also shewyd so largely your bounteousnes and liberalite anenst me that […] (1517 HC Letter by Tunstal TUNSTALL I 135)

(64) And this is very agreeable to our Doctrine, since Tersion, **besides that** it is, as I have sometimes manifestly known it, a kind or degree of Attrition, frees the Surface from those adherences that […] (1675–76 HC Boyle BOYLE 10)

(65) M. Banester […] had drawn owt a sheet of paper for to be set on the mayn mast with prayers for morning and evening and sygnes to knoe when they shold be syck which **besyde** yt was unmeasurably beyond al modesty, the conceyt was also so grosse that yf a mans head had but Aked he wold put them in fear of the frensy, the pestilent fever, […] (1582 HEC Madox Diary MADOX 130)

The ARCHER Corpus, which includes British and American texts from the second half of the 17th century to the 1990s, gives a clear picture of the gradual establishment of the distribution of the senses of the two forms.

Table 4. *Beside* and *besides* in the ARCHER Corpus

	Preposition				**Adverb**			
	beside		*besides*		*beside*		*besides*	
	Local	Abstr	Local	Abstr	Local	Abstr	Local	Abstr
1650–1699	0	2	0	29	0	2	0	20
1700–1799	3	11	0	47	1	6	0	54
1800–1899	11	2	1	34	0	0	1	29
1900–1990	23	3	0	9	0	0	0	18

The figures of occurrence in different centuries are not comparable as, in its present form, the ARCHER Corpus does not easily provide information on the amount of text material in different periods. Table 4 shows, however, one interesting and unexpected development. The increase in the use of *beside* in concrete local contexts from the nineteenth century onwards is obvious and remarkable, while *besides* becomes restricted to abstract senses of addition, contrast, etc. As a result, *beside*, after being very scantily represented in the seventeenth and eighteenth centuries becomes more frequent than *besides* in the twentieth century, thanks to its growing popularity as the indicator of local neighbourhood.

The FLOB Corpus, which consists of British English texts from the 1990s, shows that in Present-day English the shorter form has, indeed, become popular once again: there are 70 instances of *beside* (all prepositional; 65 of local use) and 44 of *besides* (14 prep., 30 adv., all of non-local use). The corresponding Present-day American English Corpus, Frown, shows roughly similar distribution, although with a somewhat more even distribution between the two forms (*beside* 67, all prep.; *besides* 51, 34 prep., 17 adv.). Even in this corpus, there is a clearcut local/non-local division between the two forms.

The massive amount of evidence in the *British National Corpus* confirms this distribution, even with a more obvious bias towards the shorter form: 5791 instances of *beside* and 2609 of *besides*. The distribution between the prepositional and adverbial uses of *besides* is the same as in FLOB, c. 2:1 (1763 as against 846). All instances of *beside* are prepositional; there is one interesting case of the use of this preposition with *of*:

(66) Richard: Where are we ⟨unclear⟩?
[…]
PS0NE: Right **beside of** Rank Xerox. (BNC KDP, Lines 3090–92)

In Present-day English the abstract non-local use of the shorter form *beside* is mainly restricted to idiomatic phrases, *beside oneself*, *beside the point*, although the senses 'in comparison to' 'except' can occasionally be found (cf. examples (5)–(6) above).

5. Concluding remarks

The development of the prepositional and adverbial uses of *beside(s)* falls nicely within the general pattern of development of a large number of adverbial connectives (prepositions, conjuncts, subordinators) in English. These forms have their roots in Old English but the actual process of grammaticalisation takes place in Middle English and is more or less completed by the sixteenth century. The development of *beside(s)* from concrete local senses to distancing and various abstract senses indicating addition, exception or denial begins in Early Middle English and increases in speed, in an S-curve pattern, in the later fourteenth and in the fifteenth century. The period of maximal syntactic and semantic variability is the beginning of the Modern English period. In later Modern English, specialisation between the senses of the two forms can be seen, with the attribution of the concrete local senses to the form *beside*. The semantic specialisation is linked with a tendency for syntactic specification, as the adverbial use becomes restricted to the longer form *besides*.

Due to its native origin, the development of *beside(s)* falls in the same group as, e.g., *before* or *until*. This pattern of grammaticalisation can be compared and contrasted to that of the borrowed connectives, such as *except*, *because* or *despite*, or of native formations emerging in Late Middle English, such as *unless* (< *on læs þe*) or *notwithstanding*.

Notes

1. For this term, see Quirk et al. (1985). Huddleston and Pullum (2002) use the term "connective adjunct".

2. In earlier studies I have discussed the grammaticalisation of *for* and *because* (1998), *rather* (1999), *according to* (2000), *except, save, unless, outtaken* (2002a), *notwithstanding, despite, in spite of* (2002b), and *to wit* (Koivisto-Alanko and Rissanen 2002). See also Kortmann's (1997:291–335) thorough discussion of the development of the English adverbial subordinators.

3. E.g. Hopper 1990, 1991; Hopper and Traugott 1993; and Traugott 1982, 1995, 1999.

4. Descriptions of these corpora can be easily found on the Internet, notably on the ICAME homepage: http://www.hd.uib.no/icame/newcd.htm. For a list and brief introductions of the historical corpora, see also Rissanen (2000b).

The acronym or abbreviated title of the corpus is given in the reference line of each example in this paper.

5. This figure is, however, affected by high frequencies in certain texts. In the medical recipes (*Leechdoms*, *Lacnunga*, *Peri Didaxeon*), for instance, *side* is referred to no less than 68 times (see example (11)). In the religious context, referring to Christ's wounded side, the word occurs 56 times in a variety of texts (example (12)).

6. In fact, there are two examples, but the other occurs in an Early Middle English manuscript, dated to c. 1225 by the *Middle English Dictionary*:

> Þe sculen nu waxen wurmes **besiden**, þeo hungrie feond, þeo þe freten [devour] wulleþ. (c1225 Body & S.[2] [Wor F.174] 69/39)

7. In ME3, the 25 instances of the sides of the astrolabe in Chaucer's *Treatise of the Astroblabe* and the frequent references to geographical areas in the histories and Mandeville's *Travels* affect the proportion.

8. I have included only those senses which seem relevant to the present discussion and reorganised and renumbered them to illustrate the chronological development. The figures and letters before the date of first occurrence, in brackets, refer to the MED categories.

9. The same kind of development can be traced with other compound connectives formed with *side*, such as *alongside*, *aside*, *inside*, *outside*, although these forms are grammaticalised later than *beside(s)*. The earliest instances of *aside* are recorded from the fourteenth century (MED; *Helsinki Corpus)*. The OED gives the earliest instances of *inside* and *outside* from the sixteenth century and of *alongside*, from the eighteenth.

10. This passage is not recorded by the MED as an occurrence of *beside(s)*. After the date of the example, the corpus or database is indicated by the acronym given in the Introduction of this paper. The references to the *Helsinki Corpus* (HC) examples also include the capitalised abbreviated title of the text. These abbreviations refer to Kytö (1996:167–230).

11. It is worth noting that in five out of the six occurrences of the postpositive use X_{dat} + *beside(s)* in Layamon's *Brut* (which is the earliest texts favouring *beside(s)*), the referent X_{dat} is a personal pronoun. The same tendency can also be seen in other Middle English texts.

12. A similar development is traceable in some other basic source concepts, such as *back*, from concrete local contexts, 'the back side of a person/building', to more abstract ones, as in the phrases *go back*, *to be taken aback*, *back up*, etc.

13. On the other hand, as the corpus is not systematically structured in regard to chronology or genre, information on the continuity of the development of the various senses and stages of grammaticalisation is less readily available.

14. Even in the Early Modern English laws and official letters included in the *Helsinki Corpus*, *beside(s)* is uncommon: 4 and 2 instances, respectively, out of a total of 159 occurrences.

Beside(s) occurs seven times in the 15th-century samples of the *Corpus of Early English Correspondence Sampler*. All instances are of the abstract use of the word.

15. The *Century of English Prose Corpus* (1680–1780) confirms the prevalence of *besides* in the 18th century: there are only 14 instances of *beside* (12 of these prepositions) as against 159 of *besides* (both preposition and adverb).

16. Cf. the corresponding developments of *according (as)*, *despite (that)*, *notwithstanding (that)*, *except (that)*, etc. (see the studies by Rissanen, quoted in fn. 2).

The OED gives examples of the subordinator *besides* (B.2b.) from 1579 on. Although antedatings to the OED are hardly worth mentioning, it may be pointed out, in passing, that the earliest instance of the subordinating *beside that* in the *Helsinki Corpus* dates from 1568:

> Wherefore I must needes dispraise the maner of our delicate Englishmen and women that drinke the Rhennish wine only for pleasure, whilst it is as yet as thicke as puddle or horsepisse. For **beside that** it giueth matter to make the [kidney] stone of, I haue knowen three within the space of one yere in high Germany that toke the falling sicknesse by drinking much newe Rhenishe wine,
>
> (1568 HC Turner, Book of Wines B4R-V)

References

Bosworth, J. and Toller, T. N. 1890. *An Anglo-Saxon Dictionary*. Oxford: Clarendon Press.

Heine, B., Claudi, U. and Hünnemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*. Chicago and London: University of Chicago Press.

Heine, B. and Kuteva, T. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hopper, P. J. 1990. "Principles of grammaticalization: Towards a diachronic typology". In *Language Typology. 1987. Systematic Balance in Language,* W. P. Lehmann (ed.), 157–170. Amsterdam and Philadelphia: John Benjamins.

Hopper, P. J. 1991. "On some principles of grammaticization". In *Approaches to Grammaticalization I*, E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam and Philadelphia: John Benjamins.

Hopper, P. J. and Traugott, E. C. 2003. *Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.

Huddleston, R. and Pullum, G.K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Koivisto-Alanko, P. and Rissanen, M. 2002. "*We give you to wit*: Semantics and grammaticalisation of the verb *wit* in the history of English". In *Variation Past and Present: VARIENG Studies on English for Terttu Nevalainen*, H. Raumolin-Brunberg, M. Nevala, A. Nurmi and M. Rissanen (eds.), 13–32. Helsinki: Société Néophilologique de Helsinki.

Kortmann, B. 1997. Adverbial Subordination: A Typology and History of Adverbial Sudordinators Based on European Languages. Berlin: Mouton de Gruyter.

Kytö, M. (comp). 1996. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts.* (3rd ed.) Helsinki: Department of English, University of Helsinki.

Mitchell, B. 1985. *Old English Syntax*. I–II. Oxford: Clarendon Press.

Österman, A. 2001. *"Where your Treasure is, There is your Heart": A Corpus-based Study of THERE Compounds and THERE/WHERE Subordinators in the History of English.* Helsinki: Société Néophilologique.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Rissanen, M. 1998. "Towards an integrated view of the development of English: notes on causal linking". In *Advances in English Historical Linguistics (1996)*. J. Fisiak and M. Krygier (eds.), 389–406. Berlin and New York: Mouton de Gruyter.

Rissanen, M. 1999. "On the adverbialization of RATHER: Surfing for historical data". In *Out of Corpora: Studies in Honour of Stig Johansson,* H. Hasselgård and S. Oksefjell (eds.), 49–59. Amsterdam and Atlanta: Rodopi.

Rissanen, M. 2000. "Paths of loan-word grammaticalisation: the case of *according to*". In *Words: Structure, Meaning, Function. A Festschrift for Dieter Kastovsky*, C. Dalton-Puffer and N. Ritt (eds.), 249–262. Berlin and New York: Mouton de Gruyter.

Rissanen, M. 2002a. "'Without except(ing) unless …': On the grammaticalisation of expressions indicating exception in English". In *Of Dyuersitie & Chaunge of Langage: Essays Presented to Manfred Görlach on the Occasion of his 65th Birthday*, K. Lenz and R. Möhlig (eds.), 77–87. Heidelberg: C. Winter.

Rissanen, M. 2002b. "On the grammaticalisation of the subordinators indicating exception". In *Text Types and Corpora: Studies in Honour of Udo Fries*, A. Fischer, G. Tottie and H.M. Lehmann (eds.). Tübingen: Gunter Narr.

Traugott, E.C. 1982. "From propositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization". In *Perspectives on Historical Linguistics*, W.P. Lehmann and Y. Malkiel (eds.). Amsterdam: Benjamins.

Traugott, E.C. 1995. "Subjectification in grammaticalisation." In *Subjectivity and Subjectivisation: Linguistic Perspectives,* D. Stein and S. Wright (eds.), 31–54. Cambridge: Cambridge University Press.

Traugott, E.C. 1999. "The rhetoric of counter-expectation in semantic change: a study in subjectification". In: *Historical Semantics and Cognition*, A. Blank and P. Koch (eds.), 177–196. Berlin: Mouton de Gruyter.

Are low-frequency complex prepositions grammaticalized?

On the limits of corpus data — and the importance of intuition*

Sebastian Hoffmann
University of Zurich

Using extensive corpus data, this paper investigates whether low-frequency complex prepositions (e.g. *by dint of*, *in conformity with*) can be interpreted as units of language structure even though they cannot have undergone the processes typical of more frequent grammaticalizing constructions. I critically evaluate previous treatments of frequency and discuss alternative modes of interpreting corpus data, e.g. by drawing a distinction between conceptual and absolute frequency as reflections of the cognitive representation of linguistic items. I also suggest that formal properties of grammaticalizing constructions may play an important role and that "grammaticalization by analogy" may account for the establishment of low-frequency units of language. Finally, the paper offers a detailed discussion of the methodological issues involved in using corpus data to study low-frequency phenomena.

1. Introduction

Over the past two decades, grammaticalization theory has had a constantly growing impact on the description of language. As a consequence, it has considerably improved our understanding of the processes involved in language change. In accordance with its functionalist approach to language, grammaticalization theory places great emphasis on the innovative power of language users (i.e. individual speakers/writers or speech communities). This is particularly apparent in studies that go beyond the simple description of structural changes as such and focus on the reasons why these changes take place. In this context, the problem-solving strategies involved in metaphorical extension (from

concrete to abstract concepts, Heine et al. 1991: 27ff.) and the role of conversational implicature have received particular attention. Many scholars have convincingly claimed that repeated pragmatic inferencing leads to the establishment of newly conventionalized meanings that are then encoded in the language. A typical example of this process can be seen in the case of the conjunction *since*, which has extended its originally temporal meaning to include the marking of causal relationships (see for example Traugott and König 1991). Another case in point is the ubiquitous textbook example of the development of *be going to* from a description of (purposive) spatial movement to a marker of futurity (see for example Heine et al. 1991 and Hopper and Traugott 1993).

In a theory which considers language use and language structure to be interrelated, frequency of occurrence is likely to play an important role: it appears intuitively obvious that a meaning generated by an originally context-bound conversational implicature can be conventionalized by frequent repetition in similar contexts of use. Indeed, there seems to be a general consensus that a relatively high discourse frequency is a prerequisite for a particular form to grammaticalize in the first place (for a different view, see Mair, this volume), although specific "threshold" frequencies have, to the best of my knowledge, never been suggested in the literature.

In the present paper, I hope to cast some light on the connection between frequency of occurrence and the process of grammaticalization. For this purpose, I will concentrate on the grammaticalization of complex prepositions in the English language. I believe, however, that my findings may prove to be relevant for grammaticalization processes in general.

According to Lehmann (1991: 501), "[t]he formation of new prepositions through the combination of a (mostly relational) noun with either a preposition or a case suffix is one of the most common grammaticalization processes in the world". Typical examples of complex prepositions are shown in (1) and (2):

(1) He sat in the chair *in front of* my desk. (BNC: A0R: 456)

(2) This labelling does not serve a public purpose *in terms of* nutrition or food safety. (BNC: ABC: 1134)

Some complex prepositions have been part of the English language for many centuries (e.g. *instead of*; see Schwenter and Traugott 1995), while others are much more recent additions. *In front of*, for example, which is one of the most common complex prepositions in Present-day English, only entered the language system in the 18th century (see Hoffmann 2001), and the most frequent complex preposition in the British National Corpus — *in terms of* —

reached full complex prepositional status only in the 20th century (Hoffmann, forthcoming).[1] Most lists of complex prepositions found in standard grammars and descriptions of English also include very rare items, e.g. *by dint of* and *in compliance with* (cf. Quirk et al. 1985: 670–71), both of which have less than 100 occurrences in the 100-million word British National Corpus. The crucial question in the context of the present paper is whether there are any grounds for regarding such exceedingly rare constructions as cases of grammaticalization or whether their low frequency precludes the kind of processes that have been observed for high-frequency items.

The methodology employed in this paper is that of corpus linguistics. Corpora have proven to be highly valuable sources for the investigation of many aspects of the language system. However, a study of low-frequency items on the basis of corpus data will invariably raise a number of methodological problems. Concepts such as statistical significance, the representativeness of a corpus, corpus size, the genre-specific use of a linguistic feature, etc. all present serious problems when the total number of occurrences of the item under consideration is extremely low. Apart from investigating the possible grammaticalization processes of low-frequency complex prepositions, the interpretation of my data will therefore also be aimed at testing the methodological limits of corpus linguistics.

2. The data

The data used in the current investigation stem from a number of synchronic and diachronic corpora. A brief description of each source is given below.

The British National Corpus (BNC)

The British National Corpus is a large corpus of modern English (almost 100 million words) which was designed to be a representative snapshot of the language as it was used on the British Isles towards the end of the 20th century. It contains about 90% written and 10% spoken data and all texts are richly annotated with a whole range of metatextual data (e.g. for written texts: text domain, age/sex of author/audience, medium of text; for spoken texts: interaction type, age/sex/social class of speaker, region where spoken text was captured). The results presented in this paper are based on the second release of the BNC (World Edition), which contains slightly less text than the original release.

For further information about the BNC, see Aston and Burnard (1998) or the official web-page at http://www.hcu.ox.ac.uk/BNC/.[2]

Newspapers

The second source of data for modern English consists of a corpus of seven British and American newspapers (full editorial content of the year 1999). It contains approximately 175 million words of running text. The British newspapers total about 80.5 million words whereas the American texts amount to approximately 95 million words. Table 1 gives an overview of the individual newspapers and their size.

Table 1. The newspaper corpus — 1999 editions

Newspaper	AE/BE	Abbreviation	No. of words
Boston Globe	AE	BOGL	36,330,276
Los Angeles Times	AE	LATM	41,011,444
USA Today	AE	USAT	17,593,021
Daily Mail	BE	TDMA	21,732,123
Mail on Sunday	BE	TMOS	6,190,539
The Times	BE	TLND	35,429,801
The Sunday Times	BE	TLNS	17,178,193
Total			175,465,397

It is important to note that the newspaper texts were not specifically compiled and annotated for use by linguists. As a consequence, they contain a number of formal inconsistencies that have to be kept in mind while interpreting the data. For example, the rather limited metatextual data available for individual news items varies from newspaper to newspaper, which consequently renders a direct comparison of text-domain specific features across newspapers somewhat more difficult. However, given the size and recentness of their contents, the newspaper texts nevertheless clearly constitute a valuable source for linguistic investigation. In the remaining part of this paper, the seven newspapers will be collectively referred to as the 'newspaper corpus'.

The Gutenberg texts

The Project Gutenberg's philosophy is "to make information, books and other materials available [for free] to the general public in forms a vast majority of the

computers, programs and people can easily read, use, quote, and search" (Di Micello 1992). This archive contains a large collection of non-copyright texts including many works written by British authors in the 17th, 18th and 19th centuries.[3] Of these, I downloaded 242 texts containing approximately 23.5 million words.

The texts were classified both on the basis of the date of birth of the author and on the publication date. However, the latter information could not be established for all texts. As a second method of classification, I distinguished between the very general categories of fiction vs. non-fiction.[4] I will be referring to my selection of the Gutenberg texts as the 'Gutenberg corpus'. Reference to the corpus will be made using the abbreviation GUT, followed by the author's surname, the date of birth of the author and the title of the work.

Other sources of diachronic data

A second source of diachronic data for the present paper is the electronically searchable database of illustrative quotations in the second edition of the *Oxford English Dictionary* (OED) on CD-ROM. Although not a corpus in the strict sense, the quotations represent a valuable source for finding additional examples of the items under consideration and for dating the introduction of new lexical items or syntactic structures into the English language.[5]

3. The grammaticalization of complex prepositions

Although there is a vast amount of recent literature on grammaticalization, complex prepositions have received comparatively little attention. One exception is Schwenter and Traugott (1995), who discuss the development of the substitutive prepositions *instead of*, *in place of*, and *in lieu of* over several centuries. In doing so, the authors place great emphasis on the influence of the discourse-pragmatic processes involved. The preposition *instead of*, for example, is shown to originate in the Old English *in stede* + genitival NP, where *stede* ('place') has a concrete locative meaning. Over time, the construction started to be used in contexts that invited an interpretation going beyond this strictly locative meaning, reflecting the subjective expectations of the speaker/writer about some other person's role or position:

> [Since it] is no more possible for one person to have exactly the same role as another than it is possible for one place to be exactly where another is […] the

> speaker implies a mismatch […] and the hearer is invited to make the construction relevant by drawing a subjective inference. Out of repeated uses the expectation (i.e. substitution) meaning was semanticized […] by context-induced reinterpretation. (Schwenter and Traugott 1995: 266)

In Present-day English, the influence of the original meaning has been further weakened, making sentences such as (3) possible:

(3) I just heard somebody knocking *instead of* ringing the bell. (BNC: KBW: 10122)

Apart from Schwenter and Traugott (1995), I know of no other study which focuses on the grammaticalization of English complex prepositions in any detail.[6] As an introduction to the discussion of low-frequency complex prepositions, which will form the bulk of this paper, I would therefore like to present a brief case study of the grammaticalization of a relatively frequent item, *in view of*. In doing so, I will give an overview of some of the important concepts of grammaticalization theory and discuss in how far they can be observed in a corpus-based diachronic study of complex prepositions.

3.1 The case of *in view of*

Sentences (4) and (5) exemplify the typical use of the construction *in view of* in the Gutenberg corpus:

(4) In this manner, we came *in view of* the entrance of a wood, through which we were to pass, at the farther side of the plain […] (GUT: Defoe [1661], *Robinson Crusoe*)

(5) They were yet barely *in view of* their mistress's house, when Oak fancied he saw the opening of a casement in one of the upper windows. (GUT: Hardy [1840], *Far from the Madding Crowd*)

In both examples, *view* is employed in its literal meaning, describing a situation in which a feature of the landscape or a building comes within range of eyesight. In Present-day English, however, a different picture emerges. Two typical instances are shown in (6) and (7):

(6) *In view of* the large number of requests made we are unable to advise you pre-departure if they cannot be met. (BNC: AM0: 1720)

(7) *In view of* your comments, I think we can safely tell the hotel what to do with their bed. (BNC: AJA: 583)

Both examples make it clear that the meaning of the complex preposition *in view of* has very little to do with an exercise of the power of vision. Rather it can be paraphrased using other complex prepositions such as *in consideration of* or *on account of* and there is a causal relationship between the complement of the preposition — the *requests* in (6) and the *comments* in (7) — and the action outlined in the following clause. This causal connection is even more apparent in (8), where *because of* or *due to* could be used instead with little or no change in meaning:

(8) *In view of* the great speed at which the bistable can operate, it would toggle on each bounce and the results would be unpredictable. (BNC: C91: 707)

This **semantic change** from concrete to more abstract meanings is a typical feature of grammaticalization. It is, however, important to stress that the observed change does not affect the noun *view* alone. Rather, the whole construction *in view of* has lost its compositional meaning and therefore requires interpretation as a single unit. As will be shown below, this development is mirrored in the structural properties of complex prepositions.

Typically, older and newer meanings may coexist for a long period of time, resulting — from a synchronic point of view — in a **layering** of meanings (Hopper 1991: 22).[7] Using diachronic data, this development can usually be traced quantitatively. Figure 1 shows the frequency per million words of *in view of* in its literal and its more abstract, complex prepositional use in the Gutenberg corpus and the BNC.

The complex prepositional use of *in view of* occurs first in texts by authors whose year of birth is later than 1800, and there is a marked rise in its frequency between the late Gutenberg texts and the BNC (2.6 vs. 15.3 instances pmw). This rise is typical of grammaticalizing constructions since the observed semantic changes enable them to be used in environments which would be incompatible with their original meaning. In other words, the construction has undergone the process of **generalization**. Figure 1 further shows that the literal meaning of *in view of* steadily decreases over the time covered by my two corpora. In Present-day English, as represented by the BNC, this use is exceedingly rare (only 9 instances in the whole corpus, or 0.1 pmw).

However, the dichotomy 'concrete vs. abstract' is not sufficient as an explanation of the semantic development undergone by *in view of* over the course of the last three centuries. Sentences (9) and (10) illustrate an important process observable in the grammaticalization of *in view of* as a complex preposition:

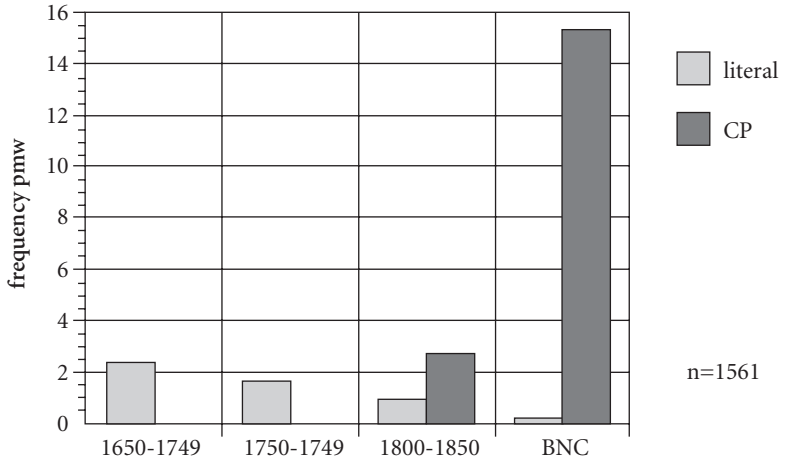


Figure 1. The distribution of literal vs. complex prepositional (CP) use of in view of in the Gutenberg corpus and the BNC.

(9) From hence we went still south about seven miles (all *in view of* this river) to Dartmouth, a town of note, seated at the mouth of the River Dart, and where it enters into the sea at a very narrow but safe entrance. (GUT: Defoe [1661], *From London to Land's End*)

(10) It is hard to see what else they could have done *in view of* the Secretary of State's decision. (BNC: CAR: 813)

In the literal meaning of *in view of* as exemplified in (9), its communicative impact is purely descriptive: it adds information about the exact topographic environment of the described scene. In the complex prepositional use of *in view of* as shown in (10), however, *in view of* has not only acquired a causal reading but furthermore expresses a subjective evaluation on the part of the author/speaker. This is a typical example of **subjectification**, which is "the development of a grammatically identifiable expression of speaker belief or speaker attitude to what is said" (Traugott 1995: 32).

In Traugott's view, the concept of subjectification offers a better framework for a description of the observed semantic changes in incipient grammaticalization than earlier approaches which centred around the process of **semantic bleaching** or **semantic attrition**, i.e. the loss of semantic content. While the original, concrete meaning of *view* certainly fades away in the grammaticalization of the complex preposition *in view of*, it is replaced by pragmatic strengthening and increased expressivity of the subjective stance of the speaker/author (cf. Hopper and Traugott 1993: 87–8; Traugott 1995: 49). Subsequent loss of

semantic content — usually going hand in hand with **phonological attrition** or **erosion** — is a typical feature of more advanced levels of grammaticalization (cf. *going to* → *gonna*).

The motivation for the observed semantic change is somewhat more difficult to trace on the basis of my diachronic data, and the following attempt at an explanation may leave many questions open. In terms of mental processes, there is an obvious connection between visual input on the one hand and cognitive evaluation of this input on the other hand: if a new (and situationally salient) object enters the visual range of a person, this may trigger a reaction. There is thus often a causal connection between vision and action. It may be the case that repeated exposure to uses of *in view of*, where the situational set-up invited **context-induced reinterpretation** towards the more subjective, causal rather than the purely visual meaning, resulted in the establishment (or **entrenchment**) of a new, conventionalized reading of the construction under consideration. As a case in point, consider sentence (11):

(11) Our conduct is *in view of* an all-seeing eye. (OED: Palmer [1710], *Proverbs*)

Here, the primary meaning of *in view of* ('within visible range') could conceivably be subject to reinterpretation because the presence of an all-seeing eye will no doubt have an influence on the conduct of the person who feels observed. A suitable paraphrase for this potentially inferred meaning would thus be 'our conduct is conditioned by the fact that there is an all-seeing eye'.

At the same time, human cognition also allows mental contemplation without ocular inspection. The potential ambiguity of sentence (11) is at least partly derived from the fact that the location of the vantage point is left open to interpretation: Is the perspective that of the all-seeing eye or is it the author's who is imagining the presence of this all-seeing eye? The faculty for introspective contemplation greatly enlarges the possible range of objects and situations available to an observer's (internal) view. Such mental contemplation is also free from the normal temporal constraints of concrete vision in that past or future events can be envisaged. What remains, however, is the fact that this type of vision may also cause a reaction on the part of its experiencer. Given that the object of the (mental) vision is not a concrete entity, the reinterpretation of the situation as subjective evaluation appears all the more likely. Unfortunately, the data for *in view of* available to me from the period of 1700–1850 is relatively sparse. In any case, a gradual transition from the visual meaning towards the modern complex prepositional use cannot be observed. A detailed study of the

discourse-pragmatic strategies at work and the connection between concrete and internal, mental vision must therefore be set aside for the time being.

From a formal point of view, the grammaticalization of complex prepositions manifests itself in a number of ways. In parallel to the semantic changes described above, the nominal element of the construction over time loses the features that define its categorial status as a noun. For example, in the complex prepositional use of *in view of*, *view* cannot occur in the plural or with a determiner, nor can it be premodified by an adjective. The noun *view* has thus undergone the process of **decategorialization**. As a result, the complex preposition *in view of* can no longer be meaningfully analysed as a fully compositional construction. Rather, the whole sequence must now be viewed holistically and treated as a single unit. It has thus undergone **syntactic reanalysis** and in terms of syntactic features now functions as the head of a prepositional phrase.

It is important to note that reanalysis is a covert process. In Langacker's classical definition, it is a "change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation" (Langacker 1977: 58). The implication is that in principle no formal distinction can be established between a (fully compositional) preposition-noun-preposition combination and a grammaticalized complex preposition. Reanalysis only becomes apparent when the use of the new structure is extended (by **analogy**) to contexts which would be incompatible with the original structure. This is for example the case when *be going to* is used with verbs that do not denote purposive meaning (e.g. *I'm going to hate this*). In the case of *in view of*, a similar situation is given when the complement of the prepositional phrase is incompatible with the concept of (ocular) vision. This is well exemplified in (12), where the abstract noun *speculation* leaves little doubt as to the complex prepositional status of the construction under consideration:

(12) The image Mr Lawson presented holed up in his home behind a securely locked gate was unfortunate, *in view of* the speculation over his political future. (BNC: A4R: 41)

There is disagreement in the literature as to whether reanalysis is a prerequisite of grammaticalization. This view seems to be taken, among others, by Hopper and Traugott (1993: 32) who claim that "[u]nquestionably, reanalysis is the most important mechanism for grammaticalization". However, the authors also stress that reanalysis cannot simply be equated with grammaticalization: "It is best [...] to regard grammaticalization as a subset of changes involved in reanalysis, rather than to identify the two" (Hopper and Traugott 1993: 50). For a radically different

opinion, see Haspelmath (1998), who claims that most grammaticalization processes can be explained without having recourse to the concept of reanalysis. In his view, explanations such as the one given for sentence (12) above are not valid since "the motivation for the various proposals of rebracketing is semantic" (Haspelmath 1998: 332). In Hoffmann (forthcoming), I offer quantitative data that support the view of complex prepositions as indivisible units which are retrieved from memory as one single chunk. While I would be wary of making any sweeping claims concerning the processes of grammaticalization in general, my findings strongly suggest that reanalysis plays an important role at least in the grammaticalization of complex prepositions. However, a detailed analysis of the interplay between reanalysis and grammaticalization must remain beyond the scope of the present investigation.

With its 1482 occurrences, *in view of* is among the 20 most frequent complex prepositions in the British National Corpus.[8] As the current section has shown, some of the processes involved in the grammaticalization of this construction can certainly be traced using diachronic corpus data. However, even with such a relatively frequent complex preposition, the quantitative analysis reached its limits. For example, it was not possible to substantiate the development from concrete to more abstract, evaluative meaning of *in view of* through potentially ambiguous instances that invite context-induced reinterpretation. In the remaining part of this paper, I will assess the value of corpus data for the study of potential grammaticalization candidates which are much less frequent than *in view of*.

4. Low-frequency complex prepositions

In order to retrieve a list of potential low-frequency complex prepositions, a frequency list of all preposition-noun-preposition (PNP) sequences occurring between 5 and 100 times in the written component of the BNC was compiled with the help of a simple Perl script. The search algorithm is illustrated in Figure 3 below. The restriction to the written component was motivated by the fact that spoken language often contains constructions which would not be compatible with commonly accepted rules of standard grammar. Consider for example the case of truncated utterances where speakers break off in the middle of a grammatical construction. In addition, hesitation phenomena, such as for example filled pauses or the (partial) repetition of individual lexical items, would have made the reliable retrieval of data more difficult.

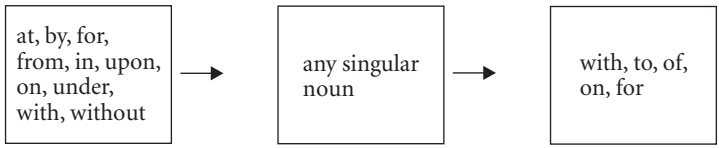


Figure 2. Retrieval algorithm for the compilation of a list of potential low-frequency complex prepositions.

Of course, the algorithm has some deficiencies. For one thing, the list of prepositions used in the search items, while certainly capturing most of the relevant forms, is not exhaustive. Also, the algorithm will not capture potential forms with plural head nouns, such as *in terms of*, or forms in which the singular noun is, either categorically or variably, preceded by an article, such as *on the strength of* or *in (the) light of*. On the other hand, there is considerable over-collection. Clearly irrelevant entries in the resulting database were deleted without consulting the corresponding sentences in the corpus. In many cases, this concerned instances where the singular noun was a place name (e.g. a city or country) as shown in (13):

(13) Rabin also attended the Socialist International meeting *in Berlin on* Sept. 15–16. (BNC: HLP: 2168)

For the remaining 1537 different types, sample sentences from the corpus were considered before I decided whether to delete them or not. In total, a list of 132 PNP-sequences was identified which were felt to be suitably parallel in usage to the established, more frequent complex prepositions. The complete list of items, including their frequency in the written component of the BNC, is shown in Table 2.

Table 2 contains a number of items (e.g. *by dint of*, *in contradistinction to*) which would also be found in a traditional enumeration of complex prepositions such as Quirk et al. (1985) and others. They have in common that the nominal element itself is a very rare lexical item which is almost exclusively encountered in the combination under investigation: *dint*, for example, occurs 68 times in the written component of the BNC, and 66 of these are instances of *by dint of*.[9] Other PNP-sequences shown in Table 2 appear to be variants of more common complex prepositions. *In face of*, for example, with 60 instances, is clearly related to *in the face of*, which occurs 1437 times in the written component of the BNC. In the case of *in search for* (13 instances), the variant part is the second prepositional element (cf. *in search of*, 972 instances in the

Table 2. Low-frequency preposition-noun-preposition sequences in the BNC (written component only)

in presence of	5	*by return of*	20	*in course of*	50
under mistake of	5	*in contemplation of*	20	*in league with*	52
in proof of	5	*by right of*	21	*under threat of*	52
in continuation of	5	*in reliance on*	21	*on pain of*	54
without breach of	5	*in liaison with*	21	*in imitation of*	54
on return of	5	*by consideration of*	23	*in sight of*	54
on publication of	5	*in appreciation of*	24	*with emphasis on*	55
in counterpoint to	5	*in furtherance of*	24	*in default of*	55
in admiration of	5	*in solidarity with*	25	*for use of*	56
in wake of	5	*for recognition of*	25	*in reaction to*	57
on review of	5	*in exercise of*	25	*for possession of*	57
in distinction to	5	*in relationship to*	25	*in wait for*	57
in want of	6	*for completion of*	25	*by analogy with*	58
on exercise of	7	*in unison with*	25	*in face of*	60
in disregard of	7	*by operation of*	25	*in retaliation for*	60
at cost of	8	*in obedience to*	26	*in concert with*	62
with relation to	8	*in discussion of*	26	*on suspicion of*	64
in emulation of	8	*on turnover of*	29	*in alliance with*	65
in token of	9	*on production of*	29	*in reference to*	65
in remembrance of	9	*by addition of*	29	*in conformity with*	65
in contradiction with	9	*for consideration of*	29	*in violation of*	66
by analogy to	9	*in expectation of*	30	*by dint of*	66
at sight of	9	*in restraint of*	30	*in sympathy with*	66
in right of	10	*for measurement of*	30	*by word of*	68
in proximity to	10	*upon receipt of*	32	*from lack of*	69
on proof of	11	*in settlement of*	33	*in praise of*	69
in commemoration of	11	*by recourse to*	34	*in tandem with*	70
in attendance on	13	*with support for*	34	*in readiness for*	72
in collusion with	13	*in compensation for*	34	*without loss of*	74
in acknowledgement of	13	*by exposure to*	35	*without fear of*	76
by contrast to	13	*by act of*	36	*in deference to*	77
in search for	13	*in collision with*	37	*by order of*	85
in hope of	14	*in support for*	39	*in company with*	88
in contradistinction to	15	*in fairness to*	40	*under cover of*	88
in hunt for	16	*in celebration of*	43	*without prejudice to*	89
in interaction with	16	*in series with*	43	*by contrast with*	90
from fear of	16	*by force of*	45	*in fear of*	90
from loss of	16	*by appeal to*	46	*in contravention of*	90
in awareness of	16	*in ignorance of*	46	*in awe of*	91
in contradiction to	17	*in respect to*	47	*at war with*	93
on entry to	18	*with news of*	47	*without regard to*	95
for love of	18	*by courtesy of*	48	*in consideration of*	97
in quest of	18	*in cooperation with*	49	*in compliance with*	98
in tribute to	19	*in point of*	50	*in harmony with*	100

written component). Many of the sequences shown in Table 2, however, are not related to more frequent complex prepositions and also do not contain very rare nominal elements (e.g. *in presence of*, *for consideration of*, and *in awareness of*). Examples (14) to (18) illustrate the typical usage of some of the items contained in Table 2. The numbers in brackets refer to the total number of instances of the PNP-construction in the written component of the BNC.

(14) *by dint of* (66):
And as every greyhaired person knows *by dint of* experience, there is no black and white issue, only grey. (BNC: B1J: 132)

(15) *in presence of* (5):
No-one after all would believe Kylie, of all people, would shed her bikini top *in presence of* the blond bombshell and frolic in the waves, unless they were an "item." (BNC: ADR: 1036)

(16) *at sight of* (9):
At sight of those silken black figures rushing their way, d'Arquebus jerked forward as though mesmerised. (BNC: CJJ: 1611)

(17) *in acknowledgement of* (13):
In acknowledgement of this, IASC, which comprises the leading world accounting bodies, sent out a separate invitation to interested parties asking for their views on the subject. (BNC: A1E: 364)

(18) *in search for* (13):
As a result it was also a place where much could be learnt and where people went, not only in need of a drink, but *in search for* something or someone. (BNC: F9U: 768)

Although all of the PNP-constructions in sentences (14) to (18) are very rare indeed, they are clearly parallel in use to much more frequent complex prepositions such as *in front of* (5915 instances) or *in relation to* (4328 instances). As a case in point, consider sentences (19) and (20):

(19) We'll tie up just beyond the lock, just *in front of* those other boats. (BNC: HHA: 1544)

(20) *In relation to* this, a number of cohesive chains were identified in the analysis, the most significant of which are the progressions of phrases relating to the cabinet ministers and the soldiers. (BNC: J89: 180)

If it is assumed that *in front of* and *in relation to* are grammaticalized sequences of words constituting units of storage in the mental representation of grammar,

is the low text-frequency of the PNP-constructions contained in sentences (14) to (18) reason enough to disallow them the same kind of complex prepositional status? In order to answer this question, I will first present an overview of relevant literature on the importance of frequency for grammaticalization (Section 5). It will become apparent that frequency of occurrence is indeed often assigned a pivotal role in the process of language change. I wish to claim, however, that an exclusive emphasis on frequency would be an oversimplification. Rather, concepts such as 'conceptual frequency' and 'saliency' can cast a different light on the role of frequency phenomena. These will be discussed in Sections 6 and 7.

It is also important to note that the perceived parallelism between high-frequency and low-frequency PNP-constructions as shown in sentences (14) to (20) relies on intuition rather than a thorough quantitative and qualitative analysis of data. In Section 8, I will present further corpus-based data and discuss their possible implications for an assessment of the grammaticalization status of low-frequency items.

5. The importance of frequency in grammaticalization

It is a basic tenet of functional linguistics that language use shapes grammar. However, the exact mechanisms which underlie the interaction between language use and language structure are a matter of continuous debate. One of the variables in the equation is certainly 'frequency of occurrence': given that a usage event is accorded the (potential) power of influencing the abstract representation of grammar in a speaker's mind, it is an obvious assumption that repeated occurrence of the same usage event will have a stronger impact than an isolated instance. Not surprisingly, then, frequency of occurrence features more or less prominently in many descriptions of grammaticalization phenomena. The current section is intended as an overview of the points raised in the relevant literature.

The direct connection between language use and language structure was already postulated towards the end of the 19th century by the German linguist Hermann Paul: "Die eigentliche Ursache für die Veränderung des Usus ist nichts anders als die gewöhnliche Sprechtätigkeit"[10] (Paul 1920: 32). In his view, language change progresses according to the same principles as those formulated by Darwin for the evolution of species: the propagation of a new form or structure depends on its usability in comparison with other competing ways of

expressing the same concept. New forms occur in the first place because speakers enjoy a certain level of freedom and creativity in language use. In order for these new forms to become accepted use, their successful application has to be negotiated between speakers and hearers, both of whose cognitive mindset ("psychischer Organismus") is influenced by this novel language use in a similar manner. Repeated application of (and exposure to) such novel usage eventually results in a shift in the language system.

Paul's emphasis on the cognitive mindset of language users which may be influenced by repeated exposure to similar usage events is reflected today in the field of cognitive grammar through Langacker's concept of entrenchment. Here, frequency of occurrence is considered to play a decisive role:

> Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; *moreover, units are variably entrenched depending on the frequency of their occurrence.* (Langacker 1987: 59, my emphasis)

Deeply entrenched items require lower processing costs in their retrieval from memory. Such an account of a gradual development towards more and more unit-like status is of course anathema to the traditional structuralist description of language with its emphasis on clearly delimited categories and constituency boundaries. It does, however, tally well with the view expressed in Pawley and Syder's influential paper (1983) in which they claim that a considerable portion of (native-like) language production relies on the application of prefabricated chunks rather than on a rule-based composition made up of individual lexical items.

The view that sequences of linguistic items can become individual units for storage in the mental lexicon through frequent usage is also often found in the literature on grammaticalization. In his discussion of routinization processes, for example, Haiman (1994) sees a link between the effect of frequent repetition (automatization), the ensuing loss of meaning via habituation, and double articulation, a fundamental property of human language:

> In the end result of automatization, of course, we can also recognize *double articulation*: the smallest meaningful signs are made up of still smaller units which are themselves meaningless. […] Sounds now meaningless may have evolved originally from meaningful morphemes. (Haiman 1994: 9, emphasis in the original)

In other words, automatization leads to reanalysis of formerly separate elements of the grammar of a language into single, meaningful units. Krug (1998) takes

a similar position, but extends the relevance of the process to other levels of grammatical structure:

> String frequency itself, on the basis of my data, can safely be considered the most important motivation in phonological and morphological changes that result in the cliticization and merger of two adjacent items across languages. From this thesis, inference to other domains seem admissible. (Krug 1998: 309)

Although Krug cautions that string frequency may not be the dominant factor in the case of the development of other syntagmatic lexical relations such as for example collocations and idioms, its influence on grammaticalization — which he views as "essentially a cognitive concept which figures on all levels of linguistic organisation" (Krug 1998: 308) — must nevertheless be seen as vital.

A more extreme position is found in Bybee (2001), who discusses the effects of frequency on French liaison, i.e. "the appearance of a word-final consonant before a vowel-initial word in words that in other contexts end in a vowel" (Bybee 2001: 338). In present-day spoken French, liaison is obligatory in a number of syntactic contexts (e.g. *vos* [*z*] *enfants*) but variable in others (e.g. *enfants* [*z*]*? intelligents* or *des découvertes* [*z*]*? inquiétantes*). In traditional accounts of the phenomenon, presence or absence of liaison is usually explained by morpho-syntactic and lexical factors as well as the degree of syntactic cohesion of the construction. Bybee argues that this approach is inadequate since it fails to consistently account for the observed variability. She claims that a more direct connection between phonology and syntax exists than is commonly assumed. Bybee shows that the likelihood for liaison to occur is highest in connection with high-frequency lexical items (e.g. *est* [*t*] *un* + noun, where virtually all instances exhibit liaison) whereas liaison is much less often found in combinations that occur more rarely. This data "supports the view that what has been called 'syntactic cohesion' is frequency of occurrence, the fact which determines the strength of the association between the first element and the second one" (Bybee 2001: 355). In other words, phonological evidence is seen as a mirror of syntactic structure. Such a view basically dispenses with many tenets of traditional syntactic theory because usage is taken as the main determining factor for constituency. Thus, highly frequent contracted forms such as *I'll* and *I'm*, which in the traditional view span across a major constituency break, would be seen as a unit.

The above-mentioned studies (Haiman 1994, Krug 1998, Bybee 2001) have in common that they see grammatical structure as a more-or-less immediate result of routinization processes which reduce the language users' cognitive

effort in encoding and decoding messages relevant to their current communicative context. In the alteration of grammatical structures, frequency of occurrence clearly is of highest importance. As Detges and Waltereit (2002: 178) pointedly state, "since routines are designed to solve some frequently occurring problem, they are, by definition, frequent themselves". Since grammaticalization is standardly defined as the process whereby content words become more grammatical (and already grammatical morphemes even more grammatical), the lexical sources for grammaticalization should consequently be recruited from frequently occurring lexical items. Indeed, as Heine et al. (1991: 32ff.) show, most of the source concepts used for grammaticalization denote fundamental elements of human experience. In terms of lexical semantics, they typically represent basic level terms (e.g. *back*, *hand*) or superordinate terms (e.g. *person*, *thing*) while "lexical items naming subordinate-level categories are not the ones likely to be grammaticalized" (Sweetser 1988: 402, quoted in Heine et al. 1991: 33). Yet, as the authors point out, it would be a mistake to overestimate the influence of frequency. Citing a study of Swahili by Bertoncini (1973), Heine et al. (1991: 39) observe that "the pool of concepts from which grammatical concepts have been drawn in Swahili is to be found in [...] approximately 20 percent of the most frequently used lexemes". However, none of the top 15 most frequent items has served as a source for grammaticalization.[11] The authors conclude that frequency alone cannot be the decisive factor in the selection of lexical items as sources for grammaticalization. Nonetheless, "it is likely to form a concomitant feature of the concepts recruited for this purpose" (Heine et al. 1991: 39).

When a lexical item (or a sequence of items) grammaticalizes, its text-frequency typically rises dramatically because the construction generalizes and can now be used in contexts which would have been incompatible with its original grammatical status and meaning (cf. the data presented for *in view of* in Section 3.1). In addition, frequent repetition leads to increased automatization and habituation, further affecting both its semantic and phonological characteristics. Once initiated, such a grammaticalization process thus proceeds in a spiralling motion (semantic bleaching, phonological attrition etc.). As a consequence, frequency can be interpreted as an indicator of the level of grammaticalization. Hopper and Traugott (1993: 103), for example, state that "[t]he more frequently a form occurs in texts, the more grammatical it is assumed to be. Frequency demonstrates a kind of generalization in use patterns".[12] However, several authors have also noted that extremely high frequency can have the opposite effect: due to its high level of entrenchment, which is

constantly being reinforced, a particular word or sequence of words may in fact resist the processes just mentioned (see for example Greenberg 1966: 68–9, Bybee 2003). In other words, high frequency can also result in what Bybee and Thompson (2000: 380) have termed the 'conserving effect': "high frequency sequences take on a life of their own, and resist change on the basis of newer productive patterns for juxtaposing words and morphemes." A typical example of this conserving effect is seen in the case of highly frequent irregular verbs (e.g. *be*, *can*, *go*, *know*, etc.) which retain their conservative irregular past tense forms rather than being replaced by the highly productive regular *-ed* pattern.

The present overview of relevant research on the role of frequency in language change in general, and grammaticalization processes in particular, has shown that frequency is generally assumed to be a highly influential variable, although comments such as those quoted from Krug (1998) make it clear that it should not be regarded as the only determining force behind the observed changes. Still, given the perceived important role of frequency, it would nevertheless seem risky to talk of the grammaticalization of low-frequency complex prepositions. In the following section, I will therefore present a number of further considerations supporting this notion.

6. Frequency and saliency

Much of the argumentation in the previous section assumes a linear connection between level of entrenchment and frequency of occurrence: the more frequently a word or a sequence of words is encountered, the more entrenched it becomes. As a case in point, consider the following quotation from Bybee:

> If we metaphorically assume that a word can be written into the [mental] lexicon, then each time a word in processing is mapped onto its lexical representation it is as though the representation was traced over again, etching it with deeper and darker lines each time. Each time a word is heard and produced it leaves a slight trace in the lexicon, it increases in lexical strength. (Bybee 1985: 117)

However, I wish to claim that such a view of the relationship between frequency of occurrence and the level of entrenchment represents an oversimplification of matters. Two important issues need to be tackled in this context: firstly, what exactly is meant by 'frequency' and secondly, what is the relationship between frequency and saliency?

In empirical linguistics, frequency of occurrence is often accorded a fairly unproblematic status. This is no doubt partly conditioned by the dramatic increase in speed and user-friendliness of computerized analyses on the basis of language corpora. Today, modern computers equipped with powerful concordancing software can be used for frequency counts of individual lexical items or whole structures in corpora containing several hundred million words. Provided that the computer is properly instructed, such an analysis is carried out reliably in a fraction of the time needed for the — much less reliable — manual scanning of printed text. However, the apparent ease with which quantitative analyses of linguistic data can be carried out is deceptive, and there are several methodological pitfalls that require careful attention.[13] One of these pitfalls, I believe, is that frequency of occurrence is given the status of an absolute variable. It is temptingly easy to discover that a feature which occurs 2000 times in one million words is more common than a feature which is on average only found 20 times in text of the same length. However, if the two features have nothing in common (e.g. meaning, functional load, etc.), the retrieved frequency information is little more than a compilation of figures. To put it differently, frequency information for an individual linguistic item only becomes meaningful as a diagnostic tool if it is compared with the frequency of occurrence of related linguistic phenomena. As a consequence, a frequency-based analysis not only needs to determine how often a particular item is actually found but also how often it *could* have occurred but in fact didn't because the concept was expressed differently. Such an approach represents a more differentiated, two-layered view of frequency phenomena. On the one hand, there is what could be termed 'conceptual frequency', i.e. the frequency with which a particular concept or speech event occurs. This type of frequency is difficult to assess by way of automated retrieval strategies because the concept can be realized in a great number of different and formally unrelated ways. On the other hand, there is also lexical frequency, i.e. how often a particular realization of the concept under investigation is found. Most importantly, however, this frequency can be expressed both in absolute terms (i.e. frequency per million words) and in terms of its relative proportion with respect to the other realizations of the same concept.

If a particular form covers the majority of realizations of a particular concept, it is the preferred choice of expression, even if its absolute frequency is relatively low. It appears likely that the cognitive representation of a linguistic item is influenced by the status it has as a preferred or dispreferred choice of expression. As a result of this, certain low-frequency items — which occur

infrequently simply because they express relatively rare events or concepts — may in fact, due to their status as preferred expressions, be more deeply entrenched than more frequent items whose use is less preferred. The decisive factor behind the level of entrenchment thus shifts away from pure frequency of occurrence to the concept of saliency (or prominence) as experienced by the language user. Such a view of the mental organization of language structure is much less mechanistic than an approach which exclusively focuses on textual frequency. It is even more sensitive to the fact that language is used and shaped in the context of social interaction rather than as an abstract and self-sufficient entity. In stating this, I do not wish to contradict the claims made by Bybee (2003) and others as presented in the previous section. In fact, I see no reason to doubt the general validity of their observations. Conceptual frequency and textual frequency of course do not exist independently of each other. If a concept is often made use of by language users, this will naturally result in a high absolute (textual) frequency of the preferred realizations of that concept. These forms will in turn be subjected to the kind of habituation processes described in Section 5. But I believe that these processes only capture certain aspects of the whole picture and that a more differentiated approach towards frequency of occurrence is required in order to cover all of its aspects and implications.

Consider for example the complex preposition *in front of*. Its meaning is closely related to that of the simple preposition *before*. In fact, in a large part of the sentences, *before* could conceivably be used instead of *in front of* without greatly altering the denotative meaning of the sentence.[14] This near-synonymous relationship is clearly shown in (21) and (22):

(21) I first saw her when I was called to see one of Mrs Ainsworth's dogs, and I looked in some surprise at the furry black creature sitting *before* the fire. (BNC: G3S: 2192)

(22) I could be comfortably at home sitting *in front of* the fire watching television. (BNC: KA6: 137)

Of course, the meaning of *before* is not restricted to the expression of a spatial relation. In fact, the temporal meaning of *before* is by far more frequent in Present-day English. From a diachronic point of view, an interesting development can be observed. Table 3 lists the frequency of both the locative and the temporal use of *before* in the early texts of the Gutenberg corpus (author birthdate 1600–1700) and the BNC. In addition, it contains the figures for *in front of*, which has an exclusively locative meaning.

Table 3. The distribution of *in front of* and *before* (in its locative and temporal meanings) in early Gutenberg texts and the BNC[a]

	Gutenberg corpus (1600–1700)		**BNC**	
	n	pmw	n	pmw
temporal *before*	1521	*1304*	78,015	*799*
locative *before*	544	*473*	7,061	*72*
in front of	0	*0*	6,123	*63*

[a]Figures for *before* are extrapolated from random subsets of 1000 sentences each.

In the early Gutenberg texts, the most frequent item is temporal *before* (1304 pmw), followed by locative *before* (473 pmw). *In front of* does not occur at all. In Present-day English, the distribution is quite different: *before* in its locative meaning has become quite rare in comparison with its temporal counterpart, which is more than ten times more frequent (72 pmw vs. 799 pmw). Most important in the present context, however, is the fact that *in front of* (63 pmw) is today almost as frequently found as locative *before* (72 pmw).

Table 3 establishes a relatively crude picture of the situation. For example, it disregards the fact that different genres and speech events will impose different stylistic restrictions on the use of the constructions available for the expression of the spatial relation under investigation. It also neglects the influence of idiomaticity on the supposed semantic equivalence of *in front of* and *before*. For example, *in front of* cannot replace *before* in *All men are equal before God*. Furthermore, there are certainly other, less frequent ways of expressing the same concept of spatial relation. A much more detailed study would indeed be required to establish the patterns of usage in a more conclusive way. These caveats clearly show that a meaningful interpretation of frequency counts is far more complex than a simple comparison of the total number of occurrences. Table 3 is nevertheless relevant for the present discussion because it shows that *in front of* has over time come to be used in a sizeable proportion of cases when this particular spatial relation needs to be expressed. Regardless of its absolute frequency, the sequence *in front of* is thus among the prime choices for the realization of this concept. If grammaticalization is indeed the result of routinization processes which reduce the language users' cognitive effort in encoding and decoding messages relevant to their current communicative context, it is highly likely that the preferred choices for a certain communicative context (e.g. the expression of a particular concept) are given a more

prominent weight in this process than less preferred choices.

For the corpus linguist, the concept of saliency is notoriously difficult to define. Since it represents a subjective assessment of familiarity, it cannot be measured in a straightforward (and quantitative) way. Some entities are thought to be salient because their physical properties make them stand out from their surrounding environment. Other items may be salient due to their cultural importance. In the second case, saliency has been linked up with 'world-frequency', i.e. the frequency with which a culturally relevant item is encountered by members of a speech community. As a result of their prominence, terms for such highly salient referents are in turn also used more frequently. However, as Croft (2000) cautions, such a view may be too simplistic:

> But what explains frequency of use? It is not frequency of occurrence in the world, but frequency of being talked about. Frequency of being talked about is a consequence of (joint) salience for the members of the relevant speech community. Hence joint salience determines frequency of use, which in terms motivates erosion of form (via the joint goal of minimizing time expended […]). (Croft 2000:75–76)

On the other hand, items or events may be salient even though they are not frequently encountered in language use. Sobkowiak (1997), for example, notes that "[s]ex, death and defecation are socio-culturally and doubtless also psychologically rather salient, but not very frequent as conversation topics". He concludes that "[t]he relationship between the different types of extralinguistic frequency, as well as between them and the (different types of) linguistic frequency may be very complex, too complex to be captured by simple deterministic claims" (ibid.).

In the present section, I have suggested that frequency can at least partly act as a suitable proxy for saliency when frequency of occurrence is seen in relation to the total number of instances of the concept expressed by the linguistic item under investigation. However, in the light of these last observations, my proposal must be subjected to further scrutiny. This impression is further supported by the fact that some of the low-frequency complex prepositions introduced in Section 4 clearly do not represent preferred choices for the concepts they express. Consider for example *in face of*, which has 60 occurrences in the written part of the BNC. As sentences (23) and (24) show, *in face of* must be considered a variant of the much more frequent combination *in the face of*, which is found 1461 times.[15]

(23) Tate Gallery perseveres with rotating the collections *in face of* mounting criticism. (BNC: CKU: 537)

(24) The problem of continued support *in the face of* declining membership is being addressed. (BNC: A67: 358)

The concepts expressed in (23) and (24) are clearly the same. The hypothesis that *in face of* represents a preferred choice for this particular concept can thus hardly be supported. If one wishes to claim that *in face of* is just as much a grammaticalized unit as its more frequent variant *in the face of*, it is therefore necessary to provide additional modes of explanation that apply independently of (the different kinds of) frequency of occurrence. In a comparable case, *in (the) light of*, preferences vary according to region, with British English preferring the form with the article, and American English the one without.

7. A tentative proposal: Grammaticalization by analogy?

In the previous sections, I have placed little emphasis on the formal characteristics of recurrent elements in language structure. Moreover, I have not been concerned with the different functional properties of such recurring elements. Rather, I have concentrated on an evaluation of the various types of frequency — i.e. conceptual, absolute textual and relative textual frequency — as a single explanatory force behind the processes underlying the grammaticalization of individual lexical items and sequences of words. Towards the end of the last section, it became clear that such a purely frequency-driven approach is too simple and that additional factors require attention. I would like to propose that one such factor is given by the shared functional and formal properties of complex prepositions.

Much research has been conducted on the nature of the prefabricated sequences of text that make up a considerable part of our everyday use of language (for an overview, see Wray 2002: 44ff.). Various types of classification have been proposed, focusing on such aspects as function, form, provenance and meaning. Consider for example the conversational routines described in Aijmer (1996) or the affective stance typically expressed by the use of idioms. It is highly likely that such interpersonal, pragmatic factors will have a profound influence on the level of entrenchment of individual sequences irrespective of their actual frequency of occurrence.[16]

While the term 'prefabricated unit' is certainly useful as a linguistic concept, it is also important to stress that it covers a wide range of phenomena. The complex prepositions under investigation in the present paper are clearly much less socio-interactional in nature than other types of recurring sequences of

language. Given their status as grammatical elements, their main task lies in the structuring of text on an informational level of organization. Regardless of their overall frequency, all of the complex prepositions under investigation have in common that they express a relation between two entities. While their individual meanings are clearly not identical, they all share an abstract, text-organising property. At the same time, the complex prepositions are structurally alike: all of them consist of two simple prepositions with an intervening nominal element. There is thus, at least on a certain level, a strong congruence between form and function. I would like to suggest that the grammaticalization of such relatively frequent complex prepositions as *in front of* and *in view of* has an impact on the grammatical status of (formally) parallel constructions that occur much more rarely. In other words, low-frequency combinations grammaticalize by analogy to their more frequent "structural relatives". In such a view of the mental organisation of language, the saliency of a particular combination of words is at least partly defined by its formal parallelism to more frequent constructions. As a consequence, very rare PNP-sequences, such as *in light of* in British usage, will be interpreted (and stored in memory) as single units even though the routinization processes normally involved in this development do not apply. As an extension to this proposal, it could also be hypothesized that certain aspects of grammaticalization rely much less on the nature and context-dependent use of individual content words than previously assumed. In such an approach, grammaticalization would result in the establishment of constructional schemas whose slots can be filled with suitable lexical items.

It is important to note that this proposal is tentative in nature. While it neatly solves the problem of low-frequency complex prepositions, its general applicability to other types of grammaticalization processes would require further analysis. The complex relationship between (low) frequency and structure certainly deserves more serious linguistic and statistical treatment.

8. Low-frequency grammaticalization phenomena and corpus data: Further methodological considerations

In the present section, I will concentrate on methodological difficulties arising in the corpus-based study of rare linguistic features such as the 132 PNP-constructions discussed above. Since, as will be shown, the value of corpora does not merely reside in the statistically significant results they may yield, they remain valuable sources of data even for such rare constructions.

Qualitative vs. quantitative analysis and the problem of statistical significance

Most studies in corpus linguistics rely on the combination of quantitative and qualitative analysis of the available data. In a qualitative approach, the main focus lies on a detailed description of the phenomena under consideration. Such a description makes it possible to capture the whole range of variation observable in an individual linguistic feature since subtle shades of meaning or minute differences in use can receive adequate attention. In the course of a qualitative analysis, low-frequency items can, in principle, receive the same kind of attention as more common linguistic features. In a quantitative analysis, on the other hand, the emphasis is on the description and classification of corpus data according to a limited set of criteria. Based on the assumption that the corpus is a representative sample of language use, the linguistic findings can be used to quantify differences in the preferred patterns of linguistic realizations across different populations (e.g. British and American English). In the words of McEnery and Wilson (1996), a quantitative analysis:

> enables one to separate the wheat from the chaff: it enables one to discover which phenomena are likely to be genuine reflections of the behaviour of a language variety and which are merely chance occurrences. In the more basic task of looking non-comparatively at a single language variety, quantitative analysis enables one to get a precise picture of the frequency and rarity of particular phenomena and hence, arguably, of their relative normality or abnormality. (McEnery and Wilson 1996: 62–3)

For the low-frequency PNP-constructions studied here, a qualitative analysis poses no methodological difficulties. It may lead to valuable pieces of individual insight, but will not contribute much towards a better assessment of the central question at issue, namely the grammatical status of such constructions. As I have shown in my discussion of *in view of* in Section 3, a quantitative analysis is much more adequate for this purpose. A qualitative approach might very well be employed to describe the use of *in view of* in different periods of language use, but only a quantitative approach will yield information about the changing frequencies of individual variants in different samples and thus help trace the development of PNP-constructions towards today's complex prepositional use.

In the interpretation of quantitative results, corpus linguistics makes extensive use of statistical methods of evaluation. Among other things, such calculations enable the researcher to assess the level of confidence with which his or her results can confirm a given hypothesis. The level of confidence, however, is

heavily dependent on frequency of occurrence. If the number of instances for individual categories is very low, the observed distribution cannot be considered statistically significant. As a consequence, the results cannot be used as a reliable basis for comparison. In the case of the set of 132 PNP-constructions at issue, this situation will inevitably occur since even their total number of occurrences is very low. In other words, even if patterns could be observed that would confirm a given hypothesis, the results would be only of limited value. The likelihood that these patterns are due to chance is simply too high.

Corpus representativeness

It is a trivial observation that the lack of data for the 132 low-frequency constructions could be remedied by using larger corpora. However, for the study of rare phenomena, overall corpus size is not necessarily the only important variable. The reliability and meaningfulness of empirical data is heavily dependent on the assumption that language corpora constitute suitable mirrors of actual language use, either in its totality or at least in a wider functional domain. The choices made by the compilers of a corpus with respect to the selection and proportional representation of different text domains consequently have a direct influence on the relevance of the linguistic results. But what constitutes a representative sample of the English language? For most members of a given population, the majority of language production will be in the realm of spoken interaction. It could therefore be argued that a representative corpus would have to consist largely of spoken data with only a small proportion of texts from a range of different written registers. Changing the perspective from production to perception, on the other hand, would make the inclusion of written material entirely defensible.

However, as Biber (1993:247) notes, there is a deeper issue of principle involved. A demographically and communicatively realistic corpus, with priority on everyday spoken interaction, would "permit summary descriptive statistics for the entire language represented by the corpus", but chances are that it would be of only limited interest for linguistic research: "[r]ather, researchers require language samples that are representative in the sense that they include the full range of linguistic variation existing in a language". As a consequence, the frequency with which a particular linguistic feature is found in a corpus may in fact be quite different from the actual frequency with which an average language user is exposed to it in his or her daily language use. This observation has implications for the application of frequency-based concepts

such as those discussed in previous sections (e.g. relative text frequency, conceptual frequency, preferred choice, etc.). While I do not wish to doubt the general validity of these concepts, it may not be possible to reliably quantify their relevance — for example for the process of language change — on the basis of actual data.

For the study of low-frequency phenomena, an additional complication in connection with corpus representativeness is established by the fact that most computerized corpora consist of text samples rather than whole texts.[17] The optimal size of such a sample is given when it adequately represents the linguistic characteristics of the text type or genre to which it belongs. However, this optimal size cannot be stated in absolute terms; rather, it heavily depends on the frequency of the linguistic features under consideration. As Biber (1993: 249) points out, "frequency counts for common linguistic features are relatively stable across 1,000 word samples, while frequency counts for rare events [...] are less stable and require longer text samples to be reliably represented". As a consequence, very infrequent phenomena will certainly not be reliably represented in any corpus which was compiled on the basis of text samples.[18] It might thus very well be the case that the individual frequencies of my 132 PNP-constructions would turn out to be quite different in another corpus of comparable size and structure. Also, certain kinds of use may be accidentally overrepresented (e.g. through untypical clustering of occurrences in a small number of text samples) while others are completely absent.

Quantifying variation

It is a generally accepted view that grammaticalization is a gradient phenomenon. As a consequence, classification of linguistic realizations according to a binary system is an impossible task. In Section 3.1, I discussed how a careful diachronic study of corpus data can reveal the grammaticalized status of the relatively frequent complex preposition *in view of*. It was for example shown that the proportion of literal uses of the nominal element *view* declined sharply over the course of the centuries while the overall frequency of the construction (with an abstract meaning of *view*) dramatically increased.

In the case of low-frequency PNP-constructions, such an approach will offer much less reliable results. In part this is due to the problems concerning representativeness discussed in the previous sub-section. More importantly, the Gutenberg corpus is roughly a quarter of the size of the BNC and its individual time-spans constitute even smaller sets of data. A meaningful diachronic study

of the development of my set of low-frequency PNP-constructions is therefore impossible simply because not enough data is available.[19]

Several scholars have identified a number of formal criteria in order to determine the relative level of grammaticalization of a particular construction. In contrast to the diachronic approach described in Section 3, such formal criteria make it possible to study grammaticalization from a synchronic point of view. Lehmann (1995 [1982]), for example, established a set of six grammaticalization parameters (integrity, paradigmaticity, paradigmatic variability, structural scope, bondedness, and syntagmatic variability). On the basis of the assumption:

> that the six parameters normally correlate and jointly constitute the grammaticalization of an item or a paradigm, we may compute its global degree of grammaticalization as a function of the six values. Items and paradigms may then be compared as to their grammaticality values. (Lehmann 1995 [1982]: 168)

However, while the parameters have the advantage of being fully formal, their quantification is nevertheless not a trivial task, as Lehmann himself observes (1995 [1982]: 160ff.).

For the present purpose, Quirk et al.'s (1985: 671–2) descriptive account of complex prepositions is therefore more suitable. The authors offer nine different criteria which help in the distinction between complex prepositions and other, freer combinations. For example, *in spite of* allows no modifications whatsoever (**in the spite of* or **in spite to* or **in spites of*) whereas *in pursuit of* alternates with *in the pursuit of* and *in hot pursuit of* etc. The authors establish a "scale of cohesiveness" with *in spite of* at one extreme — behaving in all possible ways like a single word preposition — and *on the shelf by (the door)* at the other extreme, allowing all nine types of variation. Assuming that a greater level of cohesiveness corresponds to a more advanced state of grammaticalization, Quirk et al.'s list of formal criteria offers a convenient basis for a quantitative, synchronic analysis of the grammatical status of complex prepositions.

A full analysis of all nine criteria is beyond the scope of this investigation. In the present context, it will only be possible to focus on two criteria in somewhat more detail: the ability of the nominal element of a PNP-construction to be preceded by a premodifing adjective or by a determiner. Typical instances of such forms of variation of the PNP-construction *in expectation of* are shown in (26) and (27). Sentence (25) first shows the construction without any kind of internal variation:

(25) As the dog learned the association between the sound of the bell and being fed, it salivated on hearing the bell *in expectation of* its meal. (BNC: GU8: 332)

(26) A ringed hand held a thin cigar which — as if *in* ***impatient*** *expectation of* her arrival — he stubbed in a silver tray. (BNC: H82: 2823)

(27) Anne left Germany *in* ***the*** *expectation of* seeing her family again before very long. (BNC: BNN: 333)

Almost two thirds of my set of 132 low-frequency PNP-constructions (85 types) are found at least once in the written part of the BNC with a premodifying adjective before the nominal element. However, only in four cases (*in proximity to*, *with emphasis on*, *in disregard of*, and *in contradiction to*) are the variants with a premodifying adjective more frequent than the bare PNP-construction and for the large majority of constructions, adjectival premodification is rare. A similar picture emerges for the determiners *a* and *the*: more than half of all PNP-constructions are rarely found with a determiner before the nominal element. When the results for adjectival premodification and determiners are combined, 18 invariant types can be isolated. They are listed in Table 4.

Table 4. Low-frequency PNP-constructions which do not occur with a determiner or a premodifying adjective before the nominal element (*n* = number of occurrences in the written component of the BNC)

PNP	*n*	PNP	*n*	PNP	*n*
by contrast with	90	*in default of*	55	*on proof of*	11
without prejudice to	89	*on pain of*	54	*in token of*	11
in readiness for	72	*in fairness to*	41	*in remembrance of*	9
in tandem with	70	*upon receipt of*	32	*by analogy to*	8
by dint of	66	*by contrast to*	15	*in want of*	6
in wait for	57	*in commemoration of*	11	*in distinction to*	5

While Table 4 contains a few rare types whose total number of occurrences does not exceed 15 instances in the 87.3 million words of the written component of the BNC, ten of the types are represented with at least 30 instances. It is perhaps worth noting that this set of PNP-constructions also includes an item which commonly features in lists of complex prepositions found in reference grammars and other descriptions of English: *by dint of*. At least in the case of these slightly more common PNP-constructions, the formal invariance observed in the BNC could be taken as a tentative basis for the claim that we are

dealing with indivisible units rather than sequences of words which are assembled according to the rules of syntax.[20]

The limitation of such an approach is that it disregards the phenomenon of 'layering', i.e. the fact that several stages of grammaticalization can coexist over considerable stretches of time. As a consequence, this method can contribute little to an evaluation of the grammatical status of the remaining PNP-constructions which are also found with internal variation. More importantly, it is a purely formal approach and crucial aspects such as meaning or function are completely ignored. Since grammaticalization is seen as a phenomenon which arises from the use of language in different situational contexts, a categorization on the basis of such exclusively formal aspects can thus only offer supportive evidence.

The distribution of low-frequency PNP-constructions

The frequency information for the set of 132 PNP-constructions at issue has so far been restricted to the total number of occurrences in the written component of the BNC. I have thus not taken into consideration that individual items may be unevenly distributed over the various texts (or text domains) represented in the corpora. In an extreme case, a low-frequency PNP-construction could theoretically constitute an idiosyncratic use of a single speaker. Since grammar is the result of a joint process of conventionalization, such a case could then of course not be considered relevant for the present purpose. However, none of the low-frequency items under consideration is restricted to a single text. On the contrary, the overwhelming majority exhibits an astonishingly wide distribution over different texts. A simple measure for the degree of distribution can be found in the ratio between the total number of occurrences in the written component of the BNC and the number of different texts in which the item occurs. In the case of *by dint of*, for example, this factor is 1.2 (66 instances in 55 different texts). The highest factor among my set of items is found for *in series with* (4.3, or 43 instances in 10 different texts) but this is the exception rather than the norm: only 11 PNP-constructions have a factor above 2.0. In other words, the hypothesis that many of the low-frequency items under consideration represent isolated — and possibly idiosyncratic — uses by a small set of authors can safely be discarded.

While such raw distribution data can already provide an impression of the use of low-frequency PNP-constructions, it leaves many questions unanswered. An analysis of the distribution over different text domains or genres can offer

more conclusive results. In his overview of recent grammaticalization processes in German, Lehmann (1991: 503) states that "complex prepositions arise in a narrowly circumscribed set of contexts" and only gradually spread to other areas of language use. On the basis of his data, he comments on the relationship between phraseology and grammaticalization and notes that newly formed expressions will not be perceived in isolation by other speakers but always in their specific context. Consequently, they "will not then spread at once to all kinds of contexts which, given the rules of grammar, would admit it, but will initially be restricted to certain collocations which come close to being phraseologisms" (ibid.). A wide distribution over different text domains and genres could therefore be interpreted as additional support for the claim that the PNP-constructions have indeed achieved the status of a grammatical item.

Table 5 gives an overview of the distribution of the 132 PNP-constructions over 8 different sub-corpora in the BNC which were selected on the basis of David Lee's genre classification (cf. Lee 2001).[21] The sub-corpora cover a total of 81.4 million words and contain mutually exclusive sections of the BNC. In addition, I searched the British and American components of the newspaper corpus.

As Table 5 shows, all 8 genres in the BNC contain a relatively large proportion of my low-frequency PNP-constructions. In fact, even in the smallest sub-corpora 'commerce' (3.76 million words) and 'biography' (3.53 million words), more than 50 per cent of all types can be found. With respect to the newspaper corpus, a difference between the British and the American component can be observed: while most of the constructions are attested in the British component

Table 5. Low-frequency PNP-constructions — number of types found in different corpora and genres

Corpus/genre	Number of words	Types (PNP)	Percent of all types
BNC written component	87,284,364	132	*100%*
non-academic prose	16,634,076	119	*90%*
academic prose	15,429,582	119	*90%*
miscellaneous	9,140,157	108	*82%*
newspapers	9,345,878	88	*67%*
fiction	16,194,885	84	*64%*
popular lore	7,376,391	77	*58%*
commerce	3,759,366	76	*58%*
biography	3,528,564	71	*54%*
newspaper corpus (BE)	80,530,659	121	*92%*
newspaper corpus (AE)	94,934,741	104	*79%*

(92 per cent of all types), a considerable number of types (21 per cent) are not found in the American data. Since the American component is in fact larger than the British component, this result could suggest that the distribution of these low-frequency items is to some extent a reflection of regional differences rather than of genre variation. However, the observed difference is too small to warrant far-reaching claims.

The fact that the PNP-constructions occur across different genres of course does not *per se* constitute a proof of their grammaticalized status. It simply confirms that the PNP-constructions under consideration are not restricted to certain narrow contexts. In other words, while it is clearly impossible to take the distributional characteristics shown in Table 5 as proof for the claim that we are indeed dealing with (grammaticalized) low-frequency complex prepositions, the data certainly also cannot be brought forward as counter-evidence against such a claim.

For a more comprehensive picture, the data in Table 5 needs to be combined with a closer study of individual PNP-constructions. Consider for example the case of *in readiness for*. Typical examples are shown in (28) and (29):

(28) I am now writing a history of the company *in readiness for* its centenary in 1999. (BNC: ALW: 1331)

(29) These muscles tense *in readiness for* action. (BNC: EB1: 372)

In readiness for occurs 72 times in 65 different texts of the written component of the BNC and is found in all of the 8 genres shown in Table 5. In addition to this, *in readiness for* exhibits no signs of being part of a larger sequence of words which functions as phraseologism. The complement noun phrases following *in readiness for* pertain to a whole range of contexts and situations (e.g. *slaughter*, *rematch*, *lunch-time trade*, and *the French*) and few occur more than once (e.g. *return*, with 4 instances). Also, no strong collocational bonds can be observed with respect to the verbs preceding the construction. As a matter of fact, the large majority of the 132 PNP-constructions under consideration exhibit similar combinatorial freedom.[22] Again, while this cannot be taken as conclusive proof for the complex-prepositional status of these constructions, the data available clearly does not speak against such an interpretation.

Intuition and corpus data

Over the past few decades, computerized corpora have proven to be invaluable sources for the empirical study of linguistic phenomena. In recent years, the

availability of large corpora such as the BNC has enabled an even more precise description of both language structure and language use:

> As a corpus gets larger, it does not simply show us the same data multiplied out, e.g., each item being ten times as frequent in a corpus ten times as large. Instead, the larger corpus both turns up fresh data that did not appear at all in the smaller ones and displays the previous data in steadily finer delicacy for the range and frequency of the combinations. Hosts of regularities emerge that escaped notice in smaller data sets, and would elude unguided intuition and introspection. [...] Instead of coverage, convergence, and consensus decreasing when natural language data get rewritten into a formal notation, they are now increasing when data get treated in their naturally occurring formats. (de Beaugrande 1997: 44)

As a result, few linguists will nowadays completely deny the relevance of corpora as empirical test-beds for intuition-based hypotheses about language. The recent publication of Biber et al.'s (1999) large corpus-based grammar is further testimony to the importance of corpora in language description and language teaching.

It is perhaps a telling fact that the list of complex prepositions given in Biber et al. (1999: 75) contains quite a few items whose overall frequency in the written component of the BNC is rather low. The authors include *for want of* (163), *in consequence of* (151), *in lieu of* (141), *at variance with* (137), *in light of* (116), *in compliance with* (98), and *in conformity with* (67).[23] While it is acknowledged that "it is impossible to establish a clear borderline between free combinations and complex prepositions", no specific reason is given for the inclusion of such rare combinations into their explicitly corpus-based grammar (Biber et al. 1999: 76). It must be assumed that they were included because of intuition-based considerations of normalcy.

One of the stated intentions of the present paper is to test the limits of corpus-linguistic methodology. In the present section, I have shown that the need for statistical significance and corpus representativeness indeed imposes limitations on a meaningful interpretation of quantitative data. However, it has also become clear that low-frequency data should not be entirely discarded. While it is certainly not possible to ascertain the complex prepositional status of the 132 low-frequency PNP-constructions on the basis of the method used for *in view of* in Section 3, the quantitative analysis of low-frequency data can still offer important (albeit partial) insights into the nature of the phenomena under consideration.[24] The data presented in the current section certainly supports the intuition-based hypothesis that we are indeed dealing with low-

frequency complex prepositions. In an area where intuition must necessarily play an important role, such data can thus be employed to confirm the currency of the constructions, even though the findings will not be able to be corroborated by the normal procedures of statistical analysis.

9. Conclusion

In the present paper, I have tried to answer the question whether PNP-constructions such as *by dint of*, *in readiness for*, and *in proximity to* can be considered grammaticalized units of language functioning as complex prepositions, even though their frequency of occurrence is extremely low. Within a usage-based model of language structure, such a claim is problematic because the commonly assumed mechanism of conventionalization via frequent repetition clearly cannot be at work. In my discussion of the importance of frequency for the grammaticalization of linguistic structures, I offered a different interpretation of frequency phenomena which takes into account the fact that certain concepts require to be expressed less often than others. I suggested that the relative frequency of an individual linguistic item with respect to the sum of all possible realizations of a particular concept has an influence on the cognitive representation this linguistic item receives. Thus, if a certain low-frequency construction constitutes the preferred way of expressing a concept, it is likely to receive a more prominent weight than its overall text frequency would suggest. As a consequence, the construction could be more easily reanalysed (and stored) as an indivisible unit even though it occurs quite rarely. Rather than seeing entrenchment (or lexical strength) in a purely mechanical light, such an approach focuses on the concept of saliency as context-dependent variable.

In addition to this different interpretation of frequency phenomena, I also suggested that the grammaticalization of low-frequency constructions might occur by analogy to more frequent, structurally similar sequences. Such a view of grammaticalization would imply that the saliency of a particular combination of words is at least partly defined by its formal parallelism to more frequent constructions.

In the final section of this paper, I concentrated on the data for low-frequency PNP-constructions retrieved from the written component of the BNC and the newspaper corpus. I tried to answer the question how far a limited amount of empirical evidence can contribute to an evaluation of the grammatical status of the sequences under consideration. Corpus representativeness and statistical

significance were identified as major stumbling blocks for such an undertaking, but it was nevertheless possible to identify a number of relevant tendencies. Thus, some PNP-constructions were shown to exhibit little or no variation. In addition, I interpreted the relatively even distribution of the set of 132 PNP-constructions over different texts and genres as a sign that many of them are not restricted to highly specific contexts. While none of these points can be taken as conclusive proof for the complex-prepositional status of the constructions under consideration, it seems to me that these findings still offer some important quantitative support in an area where intuition is — and continues to be — of paramount importance.

Notes

* I am greatly indebted to Miriam Locher, Joybrato Mukherjee and Gunnel Tottie for their valuable comments and constructive criticism.

1. Some scholars, however, dispute the existence of a class of 'complex prepositions' (see e.g. Seppänen et al. 1994).

2. Aston and Burnard (1998) is based on the first release of the BNC. While their general information about the structure and contents of the corpus is still relevant, some of the word-totals given for individual metatextual categories may differ considerably from what is found in the World Edition. This is due to the fact that the bibliographic and classification information was thoroughly rechecked for the second release of the BNC.

3. Further information about the Project Gutenberg can be found at their official web-page at http://promo.net/pg/.

4. 20 texts from authors born in the 17th century were not classified into fiction and non-fiction because such a classification would probably have been highly controversial. For instance, is Swift's *A Modest Proposal* to be considered fiction or non-fiction?

5. A search in the quotations of the OED sometimes reveals instances of lexical items that occur prior to the earliest occurrence given in the relevant entry of the given item.

6. For German, however, the following two studies may be of interest: Di Meola (2000) discusses complex prepositions as part of his comprehensive study of the grammaticalization of prepositions in general. Working within a synchronic approach, his focus is on the description of the structural properties of grammaticalized complex prepositions. Lehmann (1991) mentions complex prepositions among other items in his discussion of recent changes in the German language.

7. A well-known example of layering is the construction *be going to*, which can express both spatial meaning and futurity in present-day English.

8. *In terms of*, the most frequent complex preposition in the BNC, has just over 10,000 occurrences.

9. One of the two remaining instances is a pseudo-transcription of a stretch of Norfolk dialect, with *dint* representing the standard English *didn't*. The second instance occurs in a poem and has the meaning 'a mark or impression made by a blow or by pressure' (OED, *dint*, n. Sense 3).

10. In English translation: "The true cause of the change in usage is nothing else but ordinary communicative activity."

11. The same is true also for the most frequent items of individual word-classes: while there are a number of more frequent nouns that have served as source for grammaticalization, "the noun with the highest text frequency to have been grammaticalized is *mwana* 'child', which occupies the 105th position" (Heine et al. 1991:39). Bertoncini's study is based on a relatively small corpus of 40,000 words; 1,443 types occur at least five times.

12. A similar view is expressed in Greenberg (1966:69): "Frequency […] is in fact an ever present and powerful factor in the evolution of grammatical categories and thus helps in explaining the types of synchronic states actually found".

13. For an overview of these methodological considerations, see Ball (1994).

14. The stylistic differences caused by such a replacement may of course be considerable. See below.

15. This figure refers to the overall number of occurrences of the string *in the face of*. It thus also includes sentences in which it has a literal meaning as well as instances of the idiom *to fly in the face of X*. However, the large majority of instances are complex-prepositional in nature.

16. Consider, for example, the well-known fact that the idiom *kick the bucket* is very rare in actual language use but nevertheless certainly belongs to the inventory of prefabricated sequences.

17. There are of course also corpora which include complete texts, e.g. the Bank of English.

18. With 40,000 words, the target sample size for the BNC is quite large (cf. Burnard 2000:6). However, the methodological limitation in connection with low-freqeuncy items is nevertheless relevant.

19. This observation is even more relevant when the diachronic study of different text types or genres is concerned. See also Rissanen's (1989) "mystery of vanishing reliability".

20. However, the observed invariance of the 18 PNP-constructions presented in Table 4 should clearly not be overrated. In fact, with the help of a simple Google-search of the internet, counterexamples can easily be found (e.g. *in loving remembrance of*, *in apparent readiness for*). Even in the case of the largely fossilized *by dint of*, variants such as *by the dint of* or *by heavy dint of* are quite frequently attested. The information contained in Table 4 can thus only offer tentative support for the unit-like status of the PNP-constructions under consideration and this information can certainly only be meaningfully employed in conjunction with other quantitative evidence.

21. Lee's classification scheme consists of 46 categories for the written component of the BNC. My 8 categories were formed by collapsing several of the more specific sub-genres into larger super-genres (e.g. 'W_fict_drama', 'W_fict_poetry', and 'W_fict_prose' were merged to form my category 'fiction').

22. One of the exceptions is *by return of*, which strongly collocates with *post*.

23. Interestingly, the list does not contain *by dint of*, which is clearly much more fossilized than many of the included items.

24. One promising avenue of research would be to supplement low-frequency data with data received via elicitation tests (see for example the multi-method approach described in de Mönnink 2000).

References

Aijmer, K. 1996. *Conversational Routines in English*. London and New York: Longman.

Aston, G. and Burnard, L. 1998. *The BNC Handbook. Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Ball, C. N. 1994. "Automated text analysis: cautionary tales." *Literary and Linguistic Computing* 9/4: 295–302.

Bertoncini, E. 1973. "A tentative frequency list of Swahili words." *Annali dell' Istituto Orientale di Napoli* 33 (n.s. 23): 297–363.

Biber, D. 1993. "Representativeness in corpus design." *Literary and Linguistic Computing* 8/4: 243–57.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Burnard, L. 2000. *Reference Guide for the British National Corpus (World Edition)*. Available at: http://www.natcorp.ox.ac.uk/World/urg.pdf (Accessed 2 Aug. 2002).

Bybee, J. 2003. "Mechanisms of change in grammaticization: the role of frequency." In *Handbook of Historical Linguistics*, B. D. Joseph and R. D. Janda (eds.), 602–623. Oxford: Blackwell Publishers.

Bybee, J. 2001. "Frequency effects on French liaison." In *Frequency and the Emergence of Linguistic Structure*, J. Bybee and P. Hopper (eds.), 337–359. Amsterdam: John Benjamins.

Bybee, J. 1985. *Morphology: A Study of the Relationship between Meaning and Form*. Philadelphia: John Benjamins.

Bybee, J. and Thompson, S. 2000. "Three frequency effects in syntax." *Berkeley Linguistic Society* 23: 378–388.

Bybee, J., Perkins, R. and Pagliuca, W. 1994. *The Evolution of Grammar*. Chicago: University of Chicago Press.

Croft, W. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

de Beaugrande, R. 1997. *New Foundations for a Science of Text and Discourse: Cognition, Communication, and the Freedom of Access to Knowledge and Society*. Norwood, New Jersey: Ablex Publishing Corporation.

de Mönnink, I. 2000. *On the Move. The Mobility of Constituents in the English Noun Phrase: a Multi-method Approach*. Amsterdam: Rodopi.

Detges, U. and Waltereit, R. 2002. "Reanalysis vs. grammaticalization: a semantic-pragmatic account of functional change in grammar." *Zeitschrift für Sprachwissenschaft* 21/2: 151–195.

Di Meola, C. 2000. *Die Grammatikalisierung deutscher Präpositionen.* Tübingen: Stauffenburg.
Di Micello, P. 1992. "History and philosophy of Project Gutenberg." *Project Gutenberg Official Home Site.* Available at: http://promo.net/pg/history.html (Accessed 2 Aug. 2002).
Greenberg, J. 1966. *Language Universals, with Special Reference to Feature Hierarchies.* The Hague: Mouton.
Haiman, J. 1994. "Ritualization and the development of language." In *Perspectives on Grammaticalization*, W. Pagliuca (ed.), 3–28. Amsterdam: John Benjamins..
Haspelmath, M. 1998. "Does grammaticalization need reanalysis?" *Studies in Language* 22: 315–351.
Heine, B., Claudi, U. and Hünnemeyer, F. 1991. *Grammaticalization. A Conceptual Framework.* Chicago and London: University of Chicago Press.
Hoffmann, S. 2001. "In (hot) pursuit of data: complex prepositions in Late Modern English." In *New Frontiers of Corpus Research,* P. Peters, P. Collins and A. Smith (eds.), 127–146. Amsterdam: Rodopi.
Hoffmann, S. Forthcoming. *English Complex Prepositions: Theory and Use. A Corpus-based Study.* PhD Thesis. University of Zurich, Switzerland.
Hopper, P. J. 1991. "On some principles in grammaticalization." In *Approaches to Grammaticalization*, Vol. I., E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam: John Benjamins.
Hopper, P. J. and Traugott, E. C. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.
Horn, L. R. 1988. "Pragmatic theory." In *Linguistics: The Cambridge Survey*. Vol. I, F. J. Newmeyer (ed.), 113–145. Cambridge: Cambridge University Press.
Krug, Manfred. 1998. "String frequency." *Journal of English Linguistics* 26/4: 286–320.
Langacker, R. W. 1977. "Syntactic reanalysis." In *Mechanisms of Syntactic Change*, C. N. Li (ed.), 57–139. Austin: University of Texas Press.
Langacker, R. W. 1987. *Foundations of Cognitive Grammar.* Volume 1. Stanford: Stanford University Press.
Lee, D. Y. W. 2001. "Genres, registers, text types, domains, and styles: clarifying the concepts and navigating a path through the BNC jungle." *Language Learning and Technology* 5/3: 37–72.
Lehmann, C. 1991. "Grammaticalization and related changes in contemporary German." In *Approaches to Grammaticalization*, Vol. II, E. C. Traugott and B. Heine (eds.), 493–535. Amsterdam: Benjamins.
Lehmann, C. [1982] 1995. *Thoughts on Grammaticalization.* LINCOM Studies in Theoretical Linguistics. 1. Reprint. München: LINCOM Europa.
McEnery, T. and Wilson, A. 1996. *Corpus Linguistics.* Edinburgh: Edinburgh University Press.
OED: The Oxford English Dictionary. 1989. Oxford Clarendon Press, 2nd ed.
Paul, H. 1920. *Prinzipien der Sprachgeschichte.* 5th edition. Halle: Max Niemeyer.
Pawley, A. and Syder, F. H. 1983. "Two puzzles for linguistic theory: nativelike selection and nativelike fluency." In *Language and Communication*, J. C. Richards and R. W. Schmidt (eds.), 191–226. London: Longman.
Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language.* London: Longman.

Rissanen, M. 1989. "Three problems connected with the use of diachronic corpora." *ICAME Journal* 13:16–19.

Schwenter, S.A. and Traugott, E.C. 1995. "The semantic and pragmatic development of substitutive complex prepositions in English." In *Historical Pragmatics*, A.H. Jucker (ed.), 243–273. Amsterdam: Benjamins.

Seppänen, A., Bowen, R. and Trotta, J. 1994. "On the so-called complex prepositions." *Studia Anglica Posnaniensia* 29: 3–29.

Sobkowiak, W. 1997. "On the logic of markedness arguments." Paper presented at the 30th Poznan Linguistic Meeting in May 1997. Unpublished manuscript. Available at: http://elex.amu.edu.pl/~sobkow/marked.htm (Accessed 5 March 2003).

Sweetser, E.E. 1988. "Grammaticalization and semantic bleaching." *Berkeley Linguistics Society* 14: 389–405.

Traugott, E.C. 1995. "Subjectification in grammaticalisation." In *Subjectivity and Subjectivisation*, D. Stein and S. Wright (eds.), 243–273. Cambridge: Cambridge University Press.

Traugott, E.C. and König, E. 1991. "The semantics-pragmatics of grammaticalization revisited." In *Approaches to Grammaticalization*. Vol. I, E.C. Traugott and B. Heine (eds.), 189–218. Amsterdam: John Benjamins.

Wray, A. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Life after degrammaticalisation

Plural *be*[*]

Laura Wright
University of Cambridge

The present paper challenges the notion that grammaticalisation is opposed to lexicalisation, and that degrammaticalisation automatically entails lexicalisation. An equally plausible change of status of a former grammatical formative is for it to turn into a sociolinguistic marker. This point is made on the basis of the degrammaticalisation of present plural *be* in Standard English, and its subsequent reemergence as a sociolinguistic variable. Basilectal *be* was transmitted via the speech of transported London prisoners, among others, to the New World where it is now a prominent feature of African American Vernacular English.

The data for this survey are taken from the Bridewell Court Minute Books, 1559–1625. The language therein provides evidence of the speech-community from which many of the earliest Virginia indentured servants originally came.

1. Introduction

Linguists working in the grammaticalisation model have been facing increasing criticism lately, including attempts to "deconstruct" the concept definitively by breaking it up into its constituents and thus making it superfluous (see, e.g., Newmeyer 1998 or Campbell 2001). Those wishing to defend grammaticalisation as a useful explanatory concept in diachronic linguistics have been forced to pay attention to some unresolved issues which have received sketchy or insufficient treatment in the grammaticalisation literature so far. In this process of defending and refining the model, the phenomenon of degrammatic(al)isation has taken centre stage in recent theoretical debates. In particular, a widely held but pedestrian notion that grammaticalisation is somehow opposed to lexicalisation, and that any instance of degrammaticalisation would thus

automatically entail lexicalisation has been challenged. For example, the complexities of the (non)relationship between grammaticalisation and lexicalisation are discussed in great detail in several contributions to Wischer and Diewald, eds. (2002). In this volume, Lehmann argues strongly against any overlap between the two notions on grounds of principle. Another contributor, van der Auwera, while willing to grant that degrammaticalisation may involve lexicalisation in some cases, nevertheless takes pains to point out that this is by no means necessarily so and suggests a very general definition of degrammaticalisation as "the undoing of a grammatical formative into something other than a grammatical formative" (2002: 21).

Besides a lexical item, this "other" may of course plausibly involve various types of sociolinguistic or style markers, and it is my firm conviction that further empirical work on specific relevant changes is needed before we can decide what the typical case of degrammaticalisation involves, and how degrammaticalisation should be incorporated into the model. It is in this spirit that the following case study on the degrammaticalisation of present plural *be* in Standard English, and its subsequent reemergence as a sociolinguistic variable, is offered.

As is well documented, Middle English Southern third-person present tense plural *be* was rapidly replaced by Northern *are* during the 1500s. *Are* subsequently won a place in the emerging Standard dialect, but *be* continued as a minority variant in more basilectal speech, and was transmitted via the speech of transported London prisoners, amongst others, to the New World. Third-person present tense plural *be* is still very much alive in speech-communities in the Southern states of the USA, and it has been the subject of various twentieth-century surveys because — to the extent that a link can be established between it and present-day invariant or habitual uses of *be* — it is a relevant factor in the development of African American Vernacular English.

If one views Standard English as a predominantly Southern dialect, the introduction of *are* can be viewed as a simple replacement (for reasons not clearly understood) of one form by another without any change in function. Plural *be*, having lost its grammatical function, then dropped out of the Standard system altogether. However, if one views Standard English as a supra-regional amalgam dialect, comprised of features from several dialect areas (e.g. from the North, *are*, third person singular present tense indicative *-s*, the pronouns *they*, *them*, *their*; from the South, auxiliary *do*), then it is not so surprising to find that plural *be* continued to exist in the regional dialect of the South East into the seventeenth century, and still continues to the present day

in its heartland, the South West (Wakelin 1983: 8). In America, minority-variant plural *be* also survived and flourished, in that it gained the property of indexicality of ethnicity and Southernness. I shall argue that it was precisely because *be* gained a sociolinguistic property that it retained its grammatical function, albeit in a regional dialect and at a minority ratio.

The data for this survey will be taken from the c.150,000 words that I have transcribed to date from the Bridewell Court Minute Books, 1559–1647. The Court of Bridewell sentenced inmates to transportation from London to the New World from October 1607 (the Jamestown colony in Virginia was founded in May 1607). The language therein therefore provides evidence of the speech-community from which many of the earliest Virginia indentured servants originally came. The principle of selection for transcription has been presence of speech-based features — usually, the first and second person pronouns. Once a passage that is close to verbatim speech has been identified,[1] the case to which it belongs has been transcribed in its entirety, so as to provide as much context as possible. Thus, the corpus contains as much first-person narrative as has been identified so far, as well as more formal third-person court language.

The substance of my discussion will take the following order: in Section 2 I note the continuing presence of plural *be* in sixteenth and seventeenth century English in England. Then, using the Bridewell corpus, in Section 3 I show how both plural indicative and subjunctive *be* were transmitted to the Virginia Colony via the speech of transported prisoners, amongst other speakers. The data indicates that the demise of plural *be* started in the indicative mood, and that *be* lasted longest in the subjunctive (according to an Early Modern understanding of the subjunctive). In Section 4 I briefly survey some plural *be* data from present-day U.S. Southern speech, a variant now indexical of region and race. I surmise that it is in part the old Early Modern subjunctive, no longer recognised as such, that has perdured in the American South — but this remains to be tested against a corpus of present-day Southern U.S. English. Throughout, I shall be appealing to the notion of 'layering'. "Layering is the synchronic result of successive grammaticalisation of forms which contribute to the same domain" (Hopper and Traugott 1993: 124); whereby both *are* and *be* have come to co-exist in the same speech community, sharing the same grammatical functional load, but with different sociolinguistic encoding. I suggest that the pragmatic function of encoding for social class, ethnicity and region has enabled the survival of plural *be* in the Southern U.S.

2. From Middle English plural *be* to the development of Early Modern *are*

Up to the fifteenth century, the default plural indicative form of the verb *to be* in London writing was *be*, but over the following centuries, it became replaced by the Northern dialect form, *are*. This replacement of a Southern form by a Northern one was part of the process of standardisation, whereby various features spread out from their region of origin to form a new, mixed dialect that we now know as Standard English, and *are* penetrated not only the Standard dialect but also local London speech. Northern *are* is found in London writing during the Middle English period, but plural *be* remained the more common variant as late as 1500 (Lass 1992: 141) and can still be found in eighteenth-century writing (Bailey and Ross 1988: 201), and *are* did not actually become fully stabilised until the seventeenth century according to Lass (1999: 176). In a survey of the Epistle and glosses to Spenser's *Shepheardes Calendar* of 1597, Lass (1999: 176) found that plural *be* and *are* occurred in a ratio of about 2:1, with *be* preponderating slightly in negations, subordinate clauses, and after *there*. He notes that *be* lasted longest in negation, well after *are* had become the seventeenth-century default in all other constructions.

Nevalainen and Raumolin-Brunberg (1996: 57–69) surveyed the ratio of *are* to *be* in their *Corpus of Early English Correspondence* in the subperiods 1520–1550 and 1590–1620. In both periods, their corpus contains a little over 200,000 words. They found that *be* and *are* both occurred in the 1520–1550 subperiod in varying ratios according to rank, so that the nobility, for example, used *are* at a 36% ratio, the upper gentry used *are* at a 63% ratio, and merchants used *are* at a 38% ratio. By the 1590–1620 subperiod *are* usage had increased in all social ranks, with the nobility using *are* at a 93% ratio, the lower gentry at a 87% ratio and the non-gentry at 99%. So by 1620 plural indicative *be* was still in use, but as a minority variant for all writers.

Bailey and Ross (1988) surveyed some logbooks of British ships which sailed between 1631 and 1730. They found plural *are, be* and *is* variation, with *be* becoming scarcer over the period, and *is* the predominant plural form in many of the logs (1988: 201).[2]

Thus, it is not in doubt that plural *be* continued to feature as a minority variant in London writing at the point of departure of the founder generation to the New World. There may have been a syntactic constraint on its distribution, with existential *there*, negations, and subordinate clauses being the most likely sites.

3. The language of the Minutes of the Court of Governors of the Royal Hospitals of Bridewell and Bethlem, 1559–1647

In this section I look at the patterning of *be* and *are* in the Bridewell data. These records are available for consultation on microfilm in the Guildhall Library, London.[3] From 1607 until the 1640s, numbers of children and young people were taken before the Court, mostly on a charge of vagrancy, and, with their (or their parents' or masters') consent, were transported to the new colonies of Virginia from 1607, the Bermudas from 1618, St Christopher (St Kitts) from 1628, Barbados from 1632 and New England from 1643. It was expected that they would initially work as servants, and eventually become settlers themselves. After the 1640s Londoners continued to be taken to the overseas plantations from other London courts, such as the Middlesex Sessions, which transported people into the 1700s. Also, we know that a certain amount of trafficking went on, as individuals were brought before the Court of Bridewell charged with 'spiriting' and kidnapping hapless individuals to provide a workforce for the plantations. It can be assumed therefore that the transported Londoners' speech formed part of the input to the emerging new Virginia dialect, and that as they formed the early servant class, their speech would have been the kind that the arriving African slaves would have been exposed to.

Unlike Lass' survey of Spenser's *Shepheardes Calendar* of 1597, where he found that plural *be* and *are* occurred in a ratio of about 2:1, and in keeping with Nevalainen and Raumolin-Brunberg's findings for their 1590–1620 subperiod, *are* predominates in the Bridewell data, and plural *be* and plural *is* are minor variants, as shown in Table 1.

Table 1. Third person plural *are*, *be* and *is* ratios in the Minutes of the Court of Governors of the Royal Hospitals of Bridewell and Bethlem, 1559–1647

are	166	84%
be	21	11%
is	11	5%
Total tokens	198	100%

Examples:

(1) 21 December 1576 And he sayeth that Richard watwood & melcher Pelche **is** the like bawdes

(2) 21 December 1576 More he knoweth two wemen w^{ch} **be** comon bawdes

(3) 23 December 1576 Also he saieth that the said Briary and his wyfe **are** both bawdes

Examples (1)–(3) (all transcribed by a single Court Recorder, William Stackford) show the layering effect discussed by Hopper (1991:22): "Within a broad functional domain, new layers are continually emerging. As this happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers".

Lass found the most likely sites for *be* were in subordinate clauses, negative clauses and after *there*. Unfortunately, distinguishing main clauses from subordinate clauses is not a reliable exercise in the Bridewell corpus, mainly because what would have been a main clause in direct speech becomes a subordinate clause in reported speech ("she said that …"), but this is not always signalled. Only the sense of the clause allows us to reconstruct that it was reported speech, and over long passages of testimony, multiple interpretations are sometimes possible. Nor is clause analysis with regard to co-ordination and subordination always secure.[4] Much of the pre-1620 data consist of reported speech, whereas there is very little after 1620. There are no tokens of negated plural *be* in the corpus. However, existential *there* does indeed seem to act as a trigger for plural *is*, with *there is* accounting for 6/11 plural *is* tokens, and *there be* accounting for 3/21 plural *be* tokens, as contrasted with only 6/166 instances of *there are* — four of which are from the 1640s. Consider examples (4)–(8):[5]

(4) 26 June 1576 **ther be** very many in chepe side that Do kepe M^{rs} Breame And she knoweth their names Also he saieth that one M^{r} Pendar a goldesmythe a black swarffe man wth a grym visage vsed her company at the white lyon in Lumberstrete one Lucas house a clothworker And their had thuse of the body of her the said M^{rs} Breme And also many other came thether to her And she lay thear iij wekes after christmas last And he saieth that the said Lucas & his wiffe **are** bawdes

(5) 14 September 1576 She saieth **there is** a beare clark and a meater of cloth in yeldhall whoe **are** comon Doers wth whores

(6) 2 January 1576 **there is** two brethere*n* and by reporte bothe their wyves **are** whores

(7) 22 October 1577 He saieth that the p*ar*son and his wiffe **are** bothe comen bawdes and doe kepe avawtenge house of bawderye and a comen recever of harlott*es* to be deliuered in his house./ He saieth that mystris Neell is a comen bawde and she hath kept dyvers harlott*es* for he sett one

from her house to bringe her to the p*ar*sons of Islingdon./ He saieth that the same Godlye is abell to testyfye of moche more of the bawderye and whordome done in the said Harding*es* house/ He saieth that **ther be** other that came and will declare more of his lewde delynge when tyme doth sarve

(8) 23 November 1620 Elizabeth Judrith resorteth thither from ffounders Hall & lieth there when **their is** customers And shee appelleth them to such fashion as the p*er*sons desire & findeth them apparell and pay a certainty for euery sondry apparell to the Bawde & further sayeth shee hath vsed this course but one Weeke before Michaellmas last & sayeth M^res^ ffraunces Williams att the Mill Bridge on the banck side M^res^ Liste att Lambath at the kings Armes M^res^ Leuerat at Clarkenwell att the archer, M^res^ Wilkinson in Aldersgate streate att a Barbars shopp next the goulden still, M^res^ Pickering in Tiffeny Court, M^res^ Martin in S^t^ John streate att the White Porter & Basket **are** Bawdes

However, perhaps the most decisive factor overall governing *are/be* choice is mood.

Layering was fundamental in the introduction of *are* into all the old plural *be* slots, with indicative *are* preceding the introduction of subjunctive *are*, as noted by Bailey and Ross (1988: 201): "The subjunctive use of uninflected *be* remained in StE long after its use as an indicative plural became obsolete". The Early Modern subjunctive was used in all the slots in which it is found in present-day Standard English, plus those following a group of subordinating triggers, whose members included *whether*, *if*, *until*, *so as*, *although*, *unless*, *as*, *and*, *except*, *as though*, and others.[6] Contrast 8/21 plural *be* subjunctive tokens in the Bridewell corpus, with only 1/166 plural *are* subjunctive tokens (there are no subjunctive plural *is* tokens in the corpus). In the examples given under nos. (9), (10) and (11), the layering phenomenon can be seen with the single instance of subjunctive *are* just mentioned. The subordinator *whether* usually triggers the subjunctive bare form at this date (nos. 10 and 11) so the two forms, *are* and *be*, overlap a single function.

(9) 9 July 1576 she hath faire thing*es* as purses gloves knyves & other thing*es* but **whether** they **are** geven or bought she knowith not

(10) 19 December 1576 he will go to the ffeild*es* amonge those wemen that **are** Dryinge clothes & will trye them by gevinge a quarte of wyne or so*m* other thinge **whether** they **be** for hym and **yf** they **be** he poynt*es* them tyme & place for the purpose

(11) 27 May 1578 Agnes ffrenche beinge chardged by M^r^ Babham w^th^ the Judgement*es* of god And asked **whether** her former exa*mi*na*c*ions **be** trwe or not and whether she haue saied any thinge for feare or favor she sayeth that it is all merelye trwe

No. (10) also shows *if* triggering subjunctive *be*, and (11) also shows *whether* triggering subjunctive *have* (see Appendix B for further examples). It was common practice in later sixteenth century writing for such subordinators to trigger *be*, and this continued into the seventeenth century in London writing. A single Court Recorder, William Stackford, provides most of the instances of plural indicative *be* found in the corpus, as in (12).

(12) fo 130, 130v, 23 December 1576 he resort*es* to the same horse head in S^t^ kathery*n* w^th^ the same Edward*es* many tymes to whores that **be** kept there There is in white crosse streate one m^r^ Crosse kepes ijo Daughters y^t^ **be** whores very many straungers resort*es* thither Harde by m^r^ Crosses howse there dwelles one m^res^ Tarrington a bigge woma*n* a bawde & hath alwayes a whore or ijo in her howse M^res^ Jones in the Sanctuary is a bawde & hath ijo Daughters Elizabeth is maryed and Alice vnmaryed and they **be** both whores

It may be that this Court Recorder had old-fashioned, plural indicative *be* in his own idiolect, but it is probably also the case that he did not always standardise the speech of the examinants — unlike his predecessors and successors, who presumably did. Plural indicative *be* only occurs in the Bridewell Court Minute Books in passages of reported narrative (unlike plural subjunctive *be*, which also occurs in Court directives, see Appendix nos. (11), (16), (17). My guess is that this indicates that some individuals still had old-fashioned Southern plural *be* in their idiolects. The implication for Virginian speech is that the majority of transportees would have had subjunctive plural *be* in their speech, and some transportees would have had old-fashioned Southern English plural indicative *be* as a minor variant. By the point of departure, indicative plural *be* seems to have carried unfavourable social connotations for the written medium (perhaps too old-fashioned, or too colloquial and vulgar). Nonetheless, minor variants seem to have an important role in natural language as they can transmit from generation to generation over long periods of time.

As an aside, it may be of interest to note that later London speech went in the other direction and extended the scope of *are*, so that nineteenth-century Londoners had (and still have, in negative constructions at least?) *are* in the first person singular, as noted by Ellis (1889: 235). Ellis was concerned with London

pronunciation rather than morphology, and he asked his friend Prince Louis-Lucien Bonaparte, a comparative linguist who lived in London, to provide him with information about dialect in the suburbs. With regard to non-standard *are*, the Prince obtained replies from the Rector of Bushey (just outside North West London, in the greater conurbation) and the Vicar of Willesden (in North West London). They found it conceptually hard to comply with his request for information as the London dialect had become default and in their opinion only rural dialects were deemed worthy of comment — as the Rector of Bushey put it: "This place offers no opportunity of assisting your work. The inhabitants come and go, from various places, and remain but a very short time, but chiefly from London. I will not call this place a *colluvies omnium gentium*, but very much like it, and hence has no special language or dialect." Nonetheless from nearby Rickmansworth the Prince gleaned "*I be* is not much in use, *I are* is more common" from the National Schoolmaster; and from Willesden he elicited "*be* not used, but *I are* as well as *I am*".

4. Parallels with plural *be* in present-day U. S. speech

Turning to present-day studies in America, invariant *be* has been found to be present in black speech, and also in Southern white vernacular speech. Far from remaining a dormant minor variant in danger of dying out altogether, Labov (1998: 120) has claimed that "the particle *be* is the most frequent and the most salient of the AA elements in AAVE." In present-day studies, AAVE invariant *be*, both singular and plural, has been found to carry no tense information and can appear in any temporal context, although it frequently co-occurs with verbs referring to present situations, as it often has a habitual aspect (although Montgomery and Mishoe (1999: 252, fn. 4), Bailey and Maynor (1987) and others have noted that the habitual aspect function has been somewhat overstated, and it seems to be a fairly recent development).

Much empirical work in the Southern U. S. has been done by, for example, Guy Bailey and his associates, as reported in Bailey and Bassett (1986), Bailey (1993) and elsewhere. In data from 35/122 informants (8/83 white, 27/39 black) originally collected for the *Linguistic Atlas of the Gulf States*, Bailey and Bassett found twenty-four third-person plural *be* tokens (1986: 172–9). These twenty-four plural *be* tokens can be broken down as follows:

– instances of deleted underlying modals like "they be live", where Bailey and

Bassett (1986: 179) note that the context demands that there is an underlying deleted *would* (1 token)
- plural *be* + verb+*ing* forms, such as "both them horses be pulling", which Bailey (1993) has identified as a twentieth-century innovation (7 tokens)
- plural subjunctives, working with the Early Modern understanding of the subjunctive: "if they be know it [a song], I play it" (2 tokens, both triggered by *if*)
- plural *be* in other subordinate clauses (3 tokens)
- plural *be* in negative clauses (4 tokens)
- existential *there*: "there be some in those cheap hotels" (2 tokens)

That leaves just five remaining plural *be* tokens, **all** of which are preceded by *some*, as in "you see, squirrels are different colors. Some be brown, some be spotted, and some be white, you understand". Another speaker had "some of them be bad. Some of them'll bite you", and another speaker said "some of them be Johnson grass". Is this just coincidence? Moreover, several of these categories were combined, providing an even greater motivation for uninflected *be*, as in "if they don't be high enough" (both subjunctive subordinator *if* and a negator), "some of them be putting on a rows, you know" and "some people be combing me, you know" (both *some* and *be* + verb+*ing*), "well, there was some white babies be done got on your place" (both *some* and a subordinate clause), "some of them don't be an acre" (both *some* and a negator).[7] Token numbers are low, so robust conclusions cannot be drawn, but as long ago as 1988 Bailey and Ross, referring to data from Early Modern ships' logs, observed: "the [Early Modern] alternation among *are*, *is*, and *be* has analogues in both the black and white vernaculars in the Southern United States" (1988: 202), and "in the use of plural *is* and invariant *be* […] Ship English is quite similar both to BEV and white vernaculars in the lower South" (1988: 208). This observation does not seem to have provoked much further study, probably because it was overtaken by the hunt for *be2*, whereby linguists identified a habitual-aspect marker *be* in present-day AAVE speech.[8] The Bridewell data raises the possibility that the present-day reflex of Early Modern plural *be* still exists in the Southern United States. Obviously, usage nowadays is unlikely to be identical with usage of four hundred years ago, nor can it be uniform over all speech-communities in the region. But a clearer picture of the input subjunctive patterning may help to disentangle subsequent innovations from earlier relicts.

Whilst plural *be* dropped out of mainstream Standard speech as *are* encroached on its territory and the Early Modern subjunctive system was

(almost) abandoned altogether, Southern U.S. plural *be* gained the property of marking social class and region. For documentation of a parallel development, I turn to Montgomery and Mishoe (1999). Montgomery and Mishoe studied a South Carolina enclave in Horry County and its surrounding area, where speakers use not only invariant *be*, but also third person singular *bes*. Examples of *bes*, singular and plural, from Montgomery and Mishoe (1999: 246–7) are:

habitual: "He was a good boy, but bad to drink when he was young. Hell, he still *bes* bad to drink."
punctual: "Them tractor trailor trucks *bes* going too fast. They'll flip over in a minute."
stative: "He *bes* guiltier than sin. Everybody knows he done it."
negative: "The doctor says there don't *bes* nothing he can do for that baby."

Montgomery and Mishoe came to some interesting conclusions about the history of *be* in this region, noting that its origin was from England "and that finite *be* has a long history in the language, becoming a habitual verb only in the nineteenth century". "English featured two semantically identical systems ripe for reanalysis and restructuring" (Montgomery and Mishoe 1999: 271). They state that "for white speakers in the Carolinas who use them, the two paradigms appear to differ not grammatically or semantically but sociolinguistically. *Bes* and *be* are marked vernacular forms, rarely used outside informal, in-group social situations" (1999: 251). They suggest that the use of *be* and *bes* is pragmatically charged: "The primary functions of *be* and *bes* appear to be to express social identity, solidarity, and a degree of emotional intensity."

It seems that these Early Modern minor variants have come to be indexical of social attributes many years later. In Wright (2001, 2003) I invoke Lass's notion of exaptation, whereby 'redundant' material lying around a system, which once had a formal function historically, kicks into action again precisely because it acquires another, in this case sociolinguistic, function. Yesterday's grammatical particles, which are today's layered minor variants, come to be tomorrow's pragmatically-charged markers, working against the backdrop of the mainstream major variant. The presence of "two semantically identical systems ripe for reanalysis" enables the charging of sociolinguistic properties, so that the two systems are identical grammatically, but not pragmatically.

5. Summary

A consideration of the input seventeenth-century patterning of plural *be* provides an approach for analysing present-day usage in the Southern states of America. The transported prisoners from Bridewell had plural *be* in such sites as:

- subjunctive constructions after subordinators such as *whether, if, until, although, so as, and, as, except, as though, unless;*
- other subjunctive constructions, such as the optative, hortatory/mandative, and concessive subjunctives; the subjunctives of potentiality and unreality;
- after existential *there* (*is* is also found in this slot);
- other indicative constructions — a preference for subordinate clauses and negative constructions has been identified in other texts, and may underlie Bridewell usage too.

It should be noted that even as far back as Old English, subjunctive (that is, uninflected) forms are found in factual statements, and inflected forms in constructions that usually took the subjunctive (see Rissanen 1999: 227). There has always been indicative/subjunctive overlap throughout the history of English. Speakers in the various speech-communities of the Southern states of the U.S. have developed and changed this patterning over the intervening centuries, so that plural *be* has developed a habitual aspect property for many AAVE speakers, and *bes* has become marked for third person singular in Horry County and its environs and elsewhere. As noted by Montgomery and Mishoe (1999), the motivation for retaining uninflected *be* and *bes* is their connotation of localness, providing a useful way of demarcating insiderhood, as opposed to Southernness in general. Similarly AAVE's high ratios of uninflected *be* serve as a distancing mechanism from the speech of other Americans. The pragmatic function of social markedness via the plural *are/be/is* system has helped to keep plural *be* and plural *is* alive in this part of the English-speaking world.

Similar factors might have been at work in many other apparently unmotivated changes and irregularities in the grammar, for example in the reflexive pronoun, where the grammaticalisation of *self* has left us with the irregular series *myself, yourself, himself, herself, ourselves, yourselves, themselves,* which is regularised further as *myself, yourself, hisself, herself*, etc. in many nonstandard dialects. As in most other types of grammatical change, we might arrive at a deeper understanding of grammaticalisation and degrammaticalisation if we studied the processes in their social context and on the basis of substantial corpora of authentic utterances.

Notes

* My thanks to the editors Hans Lindquist and Christian Mair, and especially to Christian Mair for his time and trouble in formulating pertinent revisions, which have been incorporated.

1. See Wright (1995) for a discussion of features that bundle with speech, such as first and second person pronouns, pragmatic particles, exclamations and interjections, null subjects, *and* as both a co-ordinator and a subordinator. Wright (2003) contains a brief exposition of the main points discussed in the present paper.

2. They do not say how many *be/is/are* tokens they found.

3. For a discussion of the Court of Bridewell, the Minute Books, their contents, and the relationship of their contents to speech, see Wright (1995, 2000, 2001, 2002, 2003, in press). The Bridewell corpus is still under construction and being enlarged.

4. See Wright (1995: 95–97) for a discussion of coordination and subordination in the Bridewell data.

5. Some of these examples are also presented and briefly discussed in Wright (2003). Italics indicate abbreviated letter-graphs in the MSS.

6. See Wright (2001) for a detailed discussion of the Early Modern subjunctive system.

7. Montgomery and Mishoe (1999: 273–4) provide an appendix including 12 plural *bes* taken from the same LAGS data, but unfortunately too little context is provided for analysis. 4 tokens are preceded by *some*.

8. *Be2* is mainly used as an aspectual marker, and is viewed as an innovation of the black community only, whether derived from an underlying creole, or innovated in recent centuries. See Clarke (1999: 334–5) for a summary of habitual *be* in various dialects.

Appendix

Table 2. Tokens of plural *be* in the Bridewell Court Minute Books, 1559–1647

	Total	Indicative	Subjunctive
are	166	165	1
be	21	13	8
is	11	11	0

A. Indicative third-person plural *be*

1. 9 September 1574 And she saide vnto him naye lett me alone & I will go downe & se that all thing*e*s **be** well & come vppe againe & so she came vppe againe
2. 28 March 1575 And they bothe confesse that they **be** suer together
3. 30 April 1575 Jo*h*n hanckocke al*ia*s Jacke of the kitchi*n* saithe there resorteth to the howse of will*i*am Cooke in kentishe strete, ffraunces Cole, & Thomas Cole his Brother, & also one Thomas Smithe, who **be** very Theves
4. 21 December 1575 But the mayde at that tyme wolde saie that manie tymes there **be** those that sitt

at the Dore whist their wifes Do make them cockoldes, & so did my M^r./

5. 26 June 1576 And he saieth that ther **be** very many in chepe side that Do kepe M^rs Breame
(speaker: Richarde Rolles, servant, about 23 years old. Court Recorder: William Stackford. Plural *are* also features in Rolles' narrative: 8 tokens *are*, 1 token *be*)
6. 26 June 1576 M^rs Breame saieth that will*ia*m Mekens in fetter lane is a bawde and his wiffe also And Dothe knowe the wemen that **be** lewde almost all And Dothe carry them to strangers And he is abell to tell of all sort*es* of men & wemen that **be** lewde
(speaker: Mrs Breame, prostitute. Court Recorder: William Stackford. Plural *are* and *is* also features in Mrs Breame's narrative: 5 tokens *are*, 3 tokens *be*, 1 token *is*)
7. 26 June 1576 And he is a whoremonger And kepeth Elizabeth Cowp*er* and others and his wiffe knoweth it & also she plaieth the harlott And he knoweth it And they kepe the dore one for another while they **be** naught
(speaker: Mrs Breame, prostitute. Court Recorder: William Stackford)
8. 21 December 1576 More he knoweth two wemen w^ch **be** comon bawdes that p*ro*mysed to bringe in goldsmythes wyfe of chepe syde to one of my Lo*rd* of Oxford*es* men to the Dolphyn in the backesyde of olde ffyshestreat
(speaker: Henry Boyer, painter. Court Recorder: William Stackford. Plural *are* and *is* also features in Boyer's narrative: 6 tokens *are*, 2 tokens *be*, 2 tokens *is*)
9. 23 December 1576 he resort*es* to the same horse head in S^t kathery*n* w^th the same Edward*es* many tymes to whores that **be** kept there There is in white crosse streate one m^r Crosse kepes ijo Daughters y^t **be** whores very many straungers resort*es* thither Harde by m^r Crosses howse there dwelles one m^res Tarrington*n* a bigge woma*n* a bawde & hath alwayes a whore or ijo in her howse M^res Jones in the Sanctuary is a bawde & hath ijo Daughters Elizabeth is maryed and Alice vnmaryed and they **be** both whores Christopher Demounte had a childe by Elizabeth that is marryed the Daughter of m^res Jones
(speaker: Melcher Pelche, pimp. Court Recorder: William Stackford. Plural *are* also features in Pelche's narrative: 1 token *are*, 3 tokens *be*)
10. 22 October 1577 He saieth that ther **be** other that came and will declare more of his lewde delynge when tyme doth sarve./
(speaker: Henry Boyer. Court Recorder: William Stackford.)

B. Subjunctive third-person plural *be*

11. 6 October 1575 the Hospitalitie of the saide Jo*h*n & Jane, w^ch allureth and entyceth many yonge men to their vtter ruyne & decay, not onelie in expendinge & consumynge their good*es* & good name, but also entisinge them to suche inconvenienses that **are** & **be** abhomynable & detestable before the face of god
(Court Recorder's summary of a written statement by ffowlke Mounslowe)
12. 19 December 1576 She sayeth that watwood*es* cheife brokinge is to bringe wemen he will go to the ffeild*es* amonge those wemen that **are** Dryinge clothes & will trye them by gevinge a quarte of wyne or som*e* other thinge **whether** they **be** for hym and **yf** they **be** he poynt*es* them tyme & place for the purpose
13. 10 October 1577 And this exa*min*ant saieth that they beinge thus in the rome eu*er*ye man soe w^th his woman Grenewood tolde this exa*min*ant that she might ther see that euerye man had ther chosen his frende And it was fallen out that she must be frende to him the said Grenewoode but saied Grenewood **althoughe** they **be** mynded to doe it here yet I will not haue my sin*n*e knowen to them but I will come home to you whervpon she tolde him wher she dwelt that he might come to her
14. 27 May 1578 Agnes ffrenche beinge chardged by M^r Babham w^th the Judgement*es* of god And asked **whether** her former exa*min*ac*i*ons **be** trwe or not and whether she haue saied any thinge for feare or favor she sayeth that it is all merelye trwe

15. 10 May 1599 The said Ward and Johan his wyfe further saye that when they haue told the said Holding of the ill rule w^{ch} he kept & being spoken of by the neighbours he would answere yf they pay for what they take **yf** they **be** whores or bawds theeues or Cutpurses knaues or rogues they should be all welcome to him
16. 6 September 1628 Henr*y* Eagle Rich Bennet Ro Stevens Henry Johnson Rich Hather {*par* le watch S^{t} Pulchers Eagle va*grant* boy old prisoner should haue gone to Virginia but they would not take him warned and a shirt given him and sent away by passe./ Hather va*grant* boy lame to be sent to S^{t} Tho*mas's* Hospitall to be cured of his lame legg was taken picking a lock spared **till** others **be** found that he confesseth rob'd the house of the Lo*rd* Keeper
17. 27 May 1639 ffrancis Cobbett Sara Cobb {sent in by Constable Symond*s* Holborne for enterteyning lewd & loose company in the night they will both goe to the Barbadoes they **are** to be kept att worke **till** the **be** transported

References

Manuscript material: London, Guildhall Library, MS Minutes of the Court of Governors of Bridewell and Bethlem: Microfilm Reels MS33011/1–2, 22 April 1559–6 May 1576; MS33011/3, 7 May 1576–19 November 1579; MS33011/4, 1 February 1597/8–7 November 1604; MS33011/5, 10 November 1604–28 July 1610; MS33011/6, 26 July 1617–3 March 1626; MS33011/7, 1 March 1626–7 May 1634; MS33011/8, 21 May 1634–7 October 1642; MS33011/9, 14 October 1642–7 July 1658.

Bailey, G. 1993. "A perspective of African American English". In *American Dialect Research*, D. R. Preston (ed.), 287–318. Amsterdam/Philadelphia: Benjamins.

Bailey, G. and Bassett, M. 1986. "Invariant be in the Lower South". In *Language Variety in the South: Perspectives in Black and White,* M. B. Montgomery and G. Bailey (eds.), 158–179. Tuscaloosa: University of Alabama Press.

Bailey, G. and Maynor, N. 1987. "Decreolisation?" *Language in Society* 16: 449–474.

Bailey, G. and Ross, G. 1988. "The shape of the superstrate: morphosyntactic features of Ship English". *English World-Wide* 9/2: 193–212.

Campbell, L. 2001. "What's wrong with grammaticalization?" *Language Sciences* 23 (2/3): 113–161.

Clarke, S. 1999. "The search for origins: habitual aspect and Newfoundland Vernacular English". *Journal of English Linguistics* 27/4: 328–340.

Ellis, A. J. 1889. *On Early English Pronunciation.* London: The Philological Society.

Hopper, P. J. 1991. "On some principles of grammaticization". In Approaches to Grammaticalization, Vol. 1, E. C. Traugott and B. Heine (eds.), 17–35. Amsterdam: Benjamins.

Hopper, P. J. and Traugott, E. C. 1993. *Grammaticalization.* Cambridge: Cambridge University Press.

Labov, W. 1998. "Co-existent systems in African-American vernacular English". In *African-American English: Structure, History, and Use*, S. S. Mufwene, J. R. Rickford, G. Bailey and J. Baugh (eds.), 110–153. London and New York: Routledge.

Lass, R. 1992. "Phonology and morphology". In *The Cambridge History of the English Language, Vol II, 1066–1476*, Norman Blake (ed.), 23–155. Cambridge: Cambridge University Press.

Lass, R. (ed.). 1999. *The Cambridge History of the English Language Vol III 1476–1776.* Cambridge: Cambridge University Press.

Lehmann, C. 2002. "New reflections on grammaticalization and lexicalization". In *New Reflections on Grammaticalization,* Gabriele Diewald and Ilse Wischer (eds.), 1–18. Amsterdam: Benjamins.

Montgomery, M. and Mishoe, M. 1999. "'He bes took up with a Yankee girl and moved up there to New York': the verb *bes* in the Carolinas and its history". *American Speech* 74.3: 240–281.

Mustanoja, T. F. 1960. *A Middle English Syntax. Part I. Parts of Speech.* Mémoires de la Société Néophilologique de Helsinki, XXIII. Helsinki: Société Néophilologique.

Nevalainen, T. and Raumolin-Brunberg, H. (eds). 1996. *Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence.* Amsterdam and Georgia: Rodopi.

Rissanen, M. 1999. "Syntax". In *The Cambridge History of the English Language,* Volume 3, 1476–1776, R. Lass (ed), 187–331. Cambridge: Cambridge University Press.

Traugott, E. C. and Heine, B. 1991. *Approaches to Grammaticalization. Volume I. Focus on Theoretical and Methodological Issues.* Amsterdam: Benjamins.

Van der Auwera, J. 2002. "More thoughts on degrammaticalization." In *New Reflections on Grammaticalization*, G. Diewald and I. Wischer (eds.), 19–29. Amsterdam: Benjamins.

Wakelin, M. 1983. "The stability of English dialect boundaries". *English World-Wide* 4.1: 1–16.

Wischer, I. and Diewald, G. (eds.). 2002. *New Reflections on Grammaticalization.* Amsterdam: Benjamins.

Wright, L. 1995. "Syntactic structure of witnesses' narratives from the sixteenth-century Court Minute Books of the Royal Hospitals of Bridewell and Bedlam." *Neuphilologische Mitteilungen* 96.1: 93–105.

Wright, L. 2000. "On the construction of some Early Modern English courtroom narratives". In *English Diachronic Pragmatics,* G. Di Martino and M. Lima (eds.), 79–102. Naples: Cuen.

Wright, L. 2001. "Third person singular present tense *-s*, *-th* and zero, 1575–1648". *American Speech* 76.3: 236–258.

Wright, L. 2002. "Third person plural present-tense markers in London prisoners' depositions, 1562–1623", *American Speech* 77.3: 242–263.

Wright, L. 2003. "Eight grammatical features of Southern United States speech present in Early Modern London prison narratives." In *English in the Southern United States* (Studies in English Language), S. J. Nagle and S. L. Sanders (eds.), 36–63. Cambridge: Cambridge University Press.

Wright, L. in press. "Depositions of sixteenth and seventeenth century Londoners deported to Virginia and the Bermudas: Third person singular present tense markers". In *Transported Dialects,* R. Hickey (ed.). Cambridge: Cambridge University Press.

Subject clitics in English

A case of degrammaticalization?*

Laurel J. Brinton
University of British Columbia

As a counterexample to unidirectionality in grammaticalization, Newmeyer (1998: 270) cites the loss of second-person singular subject clitics, e.g., in *hastou* and *wiltou*, in 16th century English (Kroch et al. 1982). These forms are a common, albeit optional, feature of Middle English. Though full *thou* forms replace *-tou/-tow* clitics in Early Modern English, second-person plural enclitics, subject proclitics, and object enclitics attest to the continued viability of clisis. This paper argues that *-tou/-tow* is a reduced form, not a clitic, its disappearance being attributable to loss of a phonological rule, not decliticization. This change predates the replacement of *thou* by *you*, the non-expression of subjects in imperatives, and the spread of *do* in questions and is sudden rather than gradual.

1. Introduction

In Chapter 5 "Deconstructing grammaticalization" of his 1998 book, Frederick Newmeyer cites numerous examples of "upgradings" which he believes undermine the concept of "unidirectionality" (from lexical to grammatical or from less grammatical to more grammatical) in grammaticalization. One such example is the decliticization of subject pronouns in forms such as *hastow* or *wiltow* (restoring *hast* + *thou* or *wilt* + *thou*) after 1550 in English (Newmeyer 1998: 270–271). In respect to unidirectionality, decliticization violates what Lehmann (1995: 147ff.) calls "coalescence", a process whereby items undergoing grammaticalization become increasingly bonded, with morphological fusion often being the precursor to phonological "attrition" (reduction and loss) as well. Coalescence includes not only cliticization but also agglutination, flexion, and univerbation.

Newmeyer's example of *hastow* and *wiltow* is based on a more elaborated argument given by Kroch, Myhill, and Pintzuk in a 1982 article dealing with the spread of periphrastic *do*. Kroch et al. begin with the premise that the encliticized status of subject pronouns in the early 16th century is "uncontroversial" (1982: 287). Their argument is as follows: with subject enclisis, there is no need for *do* insertion, whose primary purpose is to distinguish subjects and objects and maintain SVO order. Subject enclisis thus accounts for the low rate of *do* in both transitive and intransitive questions with pronoun subjects (as compared with full NP subjects). The gradual increase of *do* in these structures in the later part of the 16th century "runs in tandem with a demonstrable decliticization of the subject pronoun", which becomes either a free form or a proclitic on the following verb (1982: 288).

The contention of this paper is that the facts of subject enclisis in 16th century and earlier English are rather more complex than have been suggested. Following a brief account of what is known about decliticization, I will examine subject clitics and some related structures in Middle and Early Modern English, which have received scant attention in the literature. In the remainder of the paper I will interrogate the status of these forms and their relation to grammatical developments in the Early Modern English period, concluding with some reflections on (de)grammaticalization.

As a working definition of "clitic", I will follow Trask (1993: 46):

> [A clitic is a]n item which exhibits behaviour intermediate between that of a word and that of an affix. Typically, a clitic has the phonological form of a separate word, but cannot be stressed and is obliged to occupy a particular position in the sentences in which it is phonologically bound to an adjoining word, its host.

2. Decliticization

The process of decliticization receives mention in the literature on degrammaticalization (see, e.g., Lessau 1994), but there is no fully developed description of this process. Moreover, while an increasing number of examples of degrammaticalization are being adduced — e.g., Janda (2001: 292) claims to have found 84 different examples of this phenomenon — very few of these can be understood as decliticization proper.

In principle, decliticization is the reverse, or "mirror-image", of cliticization. For Jeffers and Zwicky, decliticization occurs when clitics "emerge, or re-

emerge, as independent words" (1980:223). Joseph and Janda (1988:196) include decliticization as part of the larger process of "demorphologization", which involves "movement out of morphology into phonology or syntax"; specifically, decliticization would be the "syntacticization of a former morphological element or process" (p. 198).[1] For Lehmann (1995:18) decliticization occurs when a form which was "exclusively clitic" at a former stage acquires an autonomous use.

Clear examples of decliticization are difficult to come by. Although Campbell (1991:295) feels that instances of decliticization "while not frequent, are reasonably well known" and Newmeyer (1998:270) asserts that the process "seems to have happened repeatedly", it is instead the case that the same few examples are cited again and again. Joseph and Janda consider the two cases they relate as "representative", presumably of a more common phenomenon, though at the same time they admit that their examples are "controversial" (1988:198). The following are the least problematical cases of decliticization in the literature:

1. Campbell (1991:290–292) cites two clitics in Estonian, the question clitic *-es* and the emphatic clitic *-ep*, which, because of the loss of vowel harmony and hence of evidence of their phonological dependency, have attained independent status. However, Campbell equivocates somewhat in classifying these forms as "clitic[s]/suffix[es]" and saying that they have been either "decliticized" or "lexicalized" (see also Nevis 1986b).
2. Nevis (1986a) discusses the development of the abessive affix *-taga* in Lapp which has undergone deaffixation and reanalysis as a clitic. More importantly, in Enontekiö Lapp, *taga* has become a postposition and then an independent adverb.
3. Matsumoto (1988) instances a number of "connective particles" ("enclitics", "enclitic particles") which, as a result of discourse strategies, have detached from clause-final position and become free sentence connectives in sentence-initial position in recent Japanese. However, he notes that the free forms do not acquire the full semantics of other independent connectives but are limited to discourse/pragmatic-oriented meaning.
4. Rubino (1994) cites two diverse examples: the unbounding of the future enclitic *-to/-nto* in Ilokano (an Austronesian language) and of the enclitic *-is* 'also' in Hungarian, an independent form meaning 'indeed' that can be reduplicated.[2]

Jeffers and Zwicky (1980:223–224) argue that Indo-European relative/inter-

rogative/indefinite forms such as Latin *quis/quid*, Greek *hos*, Sanskrit *yas*, and Hittite *ku-is/ku-it* develop from the Proto-Indo-European clitics **-kwe/kwo*, **-ye/-yo*.[3] However, Lehmann (1995: 18–19) contests these examples, arguing the we cannot be certain that they were clitics in PIE, that the reconstruction may be wrong, and that the syntax is inconsistent. A second example cited by Jeffers and Zwicky (1980: 224) of unaccented finite verbs in Proto-Indo-European is likewise summarily dismissed by Lehmann, who points out that such verbs were never clitics.

Many of the examples of decliticization cited are, in fact, examples of the shift from affix to clitic or affix to independent word, but not of clitic to independent word. For instance, the example of the Irish 1st person plural ending *-mid/-muid*, which has now become an independent pronoun supplanting the inherited form *sinn* (see Bybee, Perkins, and Pagliuca 1994: 13–14), is a shift from affix > independent word, and the example of the Spanish 1st person ending *-mos* becoming the clitic *-nos* in regional dialects of Spanish (see Janda 2001: 287–288, 301) is a shift from affix > clitic. Even the oft-cited example of the development of the so-called group genitive in English (e.g., the *Queen of England's crown*) from the inflection *-s* — whatever one's analysis of this complex change and its relation to the so-called *his* genitive (e.g., *John his book*)[4] — remains a rather dubious example as, at best, it shows movement from inflection to clitic, but not from clitic to independent word (see also Norde 2001 on a comparable development in Swedish).

Thus, there would seem to be considerable vagueness in the definitions of decliticization and a paucity of unambiguous examples of the phenomenon. Moreover, there is almost no mention of the "how" or the "why" of decliticization. Once I have examined a test case from the history of English, I will return to the question of what defines decliticization and how it might operate in a specific case.

3. Subject clitics: Historical development

3.1 Sources of data

Data for this paper were collected from a number of computerized and non-computerized corpora. Because the *Helsinki Corpus of English Texts* (HC; see Kytö 1996), the best-known and broadest-based historical corpus of English, is too small to supply a sufficient number of examples of the structures I am

examining, I have supplemented the data with examples collected from the Chaucer concordance (Oizumi 1991–1992), from the on-line Middle and Modern English Collections of the University of Virginia Electronic Text Center (U of V), from the online *Oxford English Dictionary* (*OED*), from the online *Middle English Dictionary* (*MED*), from the English Verse Drama section of Chadwyck-Healey (CHVD), which proved to be an especially important source of data because of the colloquial nature of the texts, and from the British National Corpus (BNC).

While there are some occurrences of 1st and 3rd person subject clitics,[5] including some proclitics, by far the predominant form is the 2nd person enclitic. Hence, I will restrict myself to this form.

3.2 Old English

Koopman (1990: 121–125, 1997: 77–78, 89–90) finds some evidence for phonological enclitics in the Old English period, primarily with the 2nd person pronoun in inverted structures. The encliticized ending takes different forms (*-stþu, sttu, -ttu, -sþu, -tu, -stu*):

(1) a. Petrus **lufastu** me? (CP 43.3; Koopman 1997: 78)
'Peter, do you love me?'
b. þa cwæð he, hwæt **wylttu** (Mt [WsCp] 20.21; Koopman 1990: 123)
'then he said, what do you want?'

While Koopman has found over 100 examples of such clitics in the Old English concordance (Venezky and Healey 1980), he believes that the process "was probably far more widespread than orthography indicates" (1990: 122).

The *-est* 2nd person singular ending may represent an earlier (pre-Old English) process of encliticization. The inherited (Proto-Germanic) ending was **-is(i)*. Most handbooks note that the *-t* arose out of inverted structures such as **rides þu,* where a process of encliticization, assimilation, and false morphological division yielded *ridest*, the *-est* ending then being generalized even to cases of normal SV order. Rather than encliticization, Hogg (1992: 149) sees the source of /t/ — which he terms as "rather odd" — as the introduction of /t/ in inverted structures to ease the transition between /s/ and /θ/ and then reinterpretation of /t/ as part of the inflection. For Mossé (1945: 109) the influence of the very common 2nd person forms of preterite-present verbs such as *wast, scealt, ðearft* and of verbs in *mi-* such as *bist, dest, gæst* contributes as much as encliticization to the analogical development of *-t*. In any case, whether *-est* is

the result of encliticization or not, this process occurred in the pre-history of English and cannot be seen as being part of the consciousness of speakers of Early or Classical Old English.

3.3 Middle English

In Middle English, 2nd person singular subject clitics are assumed to result from a phonological process of assimilation of the initial thorn of the pronoun (*thou*) to a preceding dental (*s, t,* or *d*) with the resulting forms *-tou, -tow, -tu, -te*, etc. (see *OED*: s.v. *thou,* def. 1; *MED*: s.v. *thou*). This assimilatory process affects weak forms such as articles and demonstratives as well as 2nd person pronouns (e.g., *and þat* > *and tat, and þe* > *and te, at þe* > *atte, art þu* > *artu, wilt þu* > *wiltu*, see Mossé 1952:40). These forms, as Strang observes (1970:262), "effectively […] become enclitics".[6] Second-person singular clitics, while not particularly frequent, are most common with closed-class items, BE, CAN, DARE, DO, HAVE, MAY, OWE, SHALL, WILL, and WIT, in questions and other inverted structures. However, such clitics are rarely attached to preterite forms. To give a sense of the number of such forms, see Table 1, which compares their occurrence in three corpora.[7]

Representative examples will now be adduced. Numbers of example tokens found in the quotations of the *Middle English Dictionary* are given to suggest the relative frequency of the individual forms.[8] Encliticized forms of ARE (*artow, artou, artu, ertow, ertu*, etc.; *OED*: s.v. *be* [v.], def. A.I.1.1b) are quite common; the *MED* quotations yield approximately 119 examples.

(2) What dostow at my neighebores hous? / Is she so fair? **Artow** so amorous? (HC; Chaucer, *CT* D.WB 239–240)[9]
'What are you doing at my neighbor's house? Is she so fair? Are you so amorous?'

In contrast, encliticized forms of CAN (*canstow, canstou, canstu*) are quite uncommon (only 5 examples in the *MED* quotations).[10]

(3) Ne **canstu** me noȝt knowe? / Ihc am horn of westernesse (HC; *King Horn* 1294–1295)
'Don't you know me? I am Horn of Westerness'

Encliticized forms of DARE (*darstou, derstou*, etc.) are also not frequent; only 9 examples are found in the *MED* quotations (see (4a)); a few examples of the past tense form (*dorstestow*) also exist (see (4b)).

Table 1. 2nd person enclitics in the three Middle English corpora

Verb		Present	Past	Negative
BE	U of Virginia	42	–	–
	Chaucer	39	–	–
	HC	5	–	1 (pres)
CAN	U of Virginia	–	–	–
	Chaucer	6	–	–
	HC	2	–	–
DARE	U of V	–	–	–
	Chaucer	3	1	–
	HC	–	–	–
DO	U of V	11	–	
	Chaucer	6	–	–
	HC	4	–	–
				–
HAVE	U of V	38	4	–
	Chaucer	51	2	–
	HC	16	1	2 (pres)
MAY	U of V	22	–	–
	Chaucer	40	3	–
	HC	4	–	–
MUST	U of V	–	–	–
	Chaucer	2	–	–
	HC	1	–	–
OWE	U of V	–	1	–
	Chaucer	–	2	–
	HC	–	–	–
SHALL	U of V	49	3	–
	Chaucer	42	8	
	HC	23	–	–
WILL	U of V	23	4	1 (past)
	Chaucer	29	6	7 (pres)
				1 (past)
	HC	11	–	–
WIT	U of V	23	–	–
	Chaucer	29	3	1 (pres)
	HC	–	–	–

(U of V = University of Virginia Middle English Collection; Chaucer = Chaucer's complete works; HC = the Middle English section of the Helsinki Corpus)

(4) a. How **darstow** seyn that fals thy lady ys? (Chaucer, *TC* 5.1279)
'How dare you say that your lady is false?'
b. **Dorstestow** that I tolde in hire ere / Thi wo? (Chaucer, *TC* 1.767–768)
'Would you dare that I told your woe into her ear?'

Encliticized forms of DO (*dostow, dostou, dostu*) are somewhat more common (12 examples in the *MED* quotations) (see (5); also (2)).[11]

(5) What **dostow** her vpon me londe Wiþ outen leue? (HC; *The Romance of Sir Beues of Hamtoun* 429)
'What are you doing here upon my land without leave?'

Encliticized forms of HAVE (*hastow, hastou, hastu*) are common (over 100 examples in the *MED* quotations) (see (6a)). There are occasional past tense forms (*hadestow, haddestow, haddestou, havedestou*) (see (6b)).

(6) a. "A, false cherl," quod he, "for Goddes bones! / This **hastow** for despit doon for the nones" (HC; Chaucer, *CT* D.Sum 2153–2154)
'"Ah, false churl", said he, "for God's bones! / You have done this out of spite on purpose"'
b. **Haddestou** leued in þis londe, / Ne had Ich þoled swiche shonde (HC; *Kyng Alisaunder* 1057)
'Had you lived in this land, I would not have suffered such disgrace'

Encliticized forms of MAY (*maystow, maistow*) are relatively common (27 examples in the *MED* quotations) (see (7a)), with an occasional past tense forms (*mihtestu, myghtestow*) (see (7b)).

(7) a. And thus certes **maistow** wel seen how greet is the diligence of nature (Chaucer, *Bo* 3.pr.11.121–122)
'And thus certainly you may see well how great is the diligence of nature'
b. How **myghtestow** for rowthe me bygile? (Chaucer, *TC* 4.262)
'How were you able for pity to deceive me?

Encliticized forms of MUST (*mustow, mostu, mostou, mostow*) are rare; the *MED* yields only 5 examples (see (8)). Encliticized OUGHT is also rare (see (9)).

(8) "Thanne **mustow** graunten", quod sche, "by semblable resoun that oon and good be o same thing" (HC; Chaucer, *Bo* 3.pr.11.44–46)
'"Then you must grant", she said, "by similar reasoning that one and good are the same thing"'

(9) Wel **oughtestow** to fall, and I to dye (Chaucer, *TC* 5.545)
'Well ought you [the palace] to fall, and I to die'

Encliticized forms of SHALL (*s(c)haltou*, *s(c)haltow*, *s(c)haltu*; *OED*: s.v. *shall*, def. A.I.6) are quite common (97 examples in the *MED* quotations) (see (10a)). Some examples of the past tense form (*sholdestow*, *shuldestow*, *sholdestou*; *OED*, s.v. *shall*, def. A.I.9) can also be found (see (10b)).

(10) a. Bynethe my buttok there **shaltow** fynde / A thyng that I have hyd in pryvetee (HC; Chaucer, *CT* D.Sum 2141–2142)
'Beneath my buttock, there you shall find a thing that I have hidden in secrecy'
b. Allas, Hector, allas! why **shuldestou** deie! (*MED*; Lydgate, *Troy Book* [Aug A.4] 3.5474)
'Alas, Hector, alas! Why should you die!'

Encliticized forms of WILL are also frequent (*wiltu, wiltou, wiltow, woltou,* etc.; *OED*: s.v. *will* [v.1], def. A.6) (see (11a)). There are 55 examples in the *MED* quotations. Unlike the other forms, past tense enclitics of this verb are quite common as well (*woldestou, woldestow, wostou; OED*: s.v. *will* [v.1], def. 9); the *MED* yields 11 examples (see (11b)).

(11) a. And seyde, "**Wiltu** ben wit me? / Gladlike wile Ich fede þe" (HC; *Havelok* 906–907)
'And said, "Will you be with me? I will gladly feed you"'
b. yif thou haddest entred in the path of this lif a voyde weyfarynge man, thanne **woldestow** syngen byfor the theef (*MED*; Chaucer *Bo* 2.pr.5.179–181)
'if you had entered in the path of this life a destitute wayfaring man, then you would sing before the thief'

Encliticized forms of WIT (*wastu, wostu, wostow, wastou, woostu,* etc.; *OED*: s.v. *wit* [v. 1], def. A.2b) also occur uncommonly in the *MED* quotations (33 examples) (see (12a)). A past tense form (*wistestow*) is also occasionally found (see (12b)).

(12) a. **Wostow** nat wel the olde clerkes sawe / That "who shal yeve a lovere any lawe"? (*MED*; Chaucer *CT* A.Kn 1163–1164)
'Do you not know well the old clerks' saying that "whoever shall give a lover any law"?'

b. For **wystestow** myn herte wel, Pandare, / God woot, of this thow woldest litel care (Chaucer, *TC* 3.1644–1645)
'For knew you my heart well, Pandarus, God knows, you would little care of this'

In addition to these closed-set items, a number of common verbs also occur with 2nd person enclitics, especially SAY (*saistow, seistow, saistou, saystu, sedestow*), SEE (*seestow*),[12] and KNOW (*knowestow*) (see Appendix (1)). Chaucer's works also contain a few examples of enclitics with ASK (4 examples), DEEM (6 examples), DESIRE (2 examples), HEAR (4 examples), HOLD (3 examples), LIE (2 examples), MAKE (3 examples), REMEMBER (2 examples), SPEAK (2 examples), TAKE (2 examples), TROW (6 examples), UNDERSTAND (3 examples), and WEEN (11 examples) (see Appendix (2)). A particularly rich example is given in (13).

(13) "**Scornestow** me", quod I, "or elles, **pleyestow** or **disseyvistow** me" (Chaucer, *Bo* 3.p.12.154–155)
'"Do you scorn me", I said, "or else, do you tease or deceive me?"'

A number of other encliticized forms — see Table 2 — occur once each in Chaucer's work. *Boethius*, which is constructed in the form of a dialogue, provides a large number of examples, as do the dialogue portions of the *Canterbury Tales*.

Doubly clitic forms, with a negative proclitic and subject enclitic, such as *nartou* < *ne art thou* (see (14a)), *nastou* < *ne hast thou* (see (14b)), *neltou* < *ne wilt thou* (see (14c)), or *noldestow* < *ne woldest thou* (see (14d)), can also be found:

Table 2. 2nd person enclitics that appear only once in the Chaucer corpus

bigilestow 'beguile'	blamestow 'blame'	comestow 'come'	cridestow 'cried'
denyestow 'deny'	desputestow 'dispute'	disseyvistow 'deceive'	dowtestow 'doubt'
drawestow 'draw'	enforcestow 'enforce'	feffedestow 'endow'	felistow 'feel'
gabbestow 'gossip'	grauntestow 'graunt'	herdestow 'heard'	hydestow 'hide'
intendestow 'intend'	juggestow 'judge'	leevestow 'leave'	lyvestow 'live'
menestow 'mean'	pleyestow 'play'	pleynestow 'complain'	prechestow 'preach'
profestow 'profess'	refusestow 'refuse'	rydestow 'ride'	schewedestow 'showed'
scornestow 'scorn'	sekestow 'seek'	slepestow 'sleep'	slombrestow 'slumber'
sorwestow 'grieve'	spilestow 'kill'	suffrestow 'suffer'	syngestow 'sing'
wentestow 'went'	wepistow 'weep'	werestow 'wear'	wroghtestow 'work'

(14) a. "**Nartou** so hardy" quoþ Candace (HC; *Kyng Alisaunder* 7752)
'"You are not so hardy", said Candace'
b. **Nastow**, sire, me fader slawe? / Þow schelt ben hanged & to-drawe,
Be Godes wille! (HC; *The Romance of Sir Beuis of Hamtoun* 433–434)
'Haven't you, sir, my father slain? You shall be hanged and drawn,
by God's will'
c. But oones **nyltow**, for thy coward herte … tellen of thy sorwes
smerte? (Chaucer, *TC* 1.792–794)
'But at once won't you, because of your cowardly heart, tell of your
painful sorrows?'
d. For **noldestow** of bownte hem socouren (Chaucer, *TC* 3.1264)
'for you would not out of goodness aid them'

The use of enclitic forms, although clearly favored, especially by Chaucer, is not obligatory. Some examples of non-encliticized forms[13] can be found in the 14th century with both closed-class items (see Appendix (3); cf. (2)–(12)) and open-class items (see Appendix 4; cf. Appendix (1)–(2), including negative enclitics (see Appendix 5; cf. (14)). By the 15th century non-encliticized forms are quite common.

Further evidence for the optionality of the cliticized forms is the V + *nat/not* + *thou* structure in which the negative intercedes between the verb and the subject.

(15) a. "**Art nat thou** he," quod sche, "that whilom, norissched with my
melk?" (*MED*; Chaucer, *Bo* 1.pr.2.4–6)
'"Are you not he", she said, "that once nourished with my milk?"'
b. **Nost not thou** / That ys betyd, lo, late or now? (*MED*; Chaucer, *HF*
2048)
'Do you not know what has happened, lo, just recently?'

Kroch et al. (1982: 289–290) argue that insertion of *not* between the inverted verb and pronominal subject after 1550 points to decliticization of the subject pronoun. Though rare, such forms clearly occur earlier than Kroch et al. claim. Rissanen (1999a) finds 5 examples of this order (Auxiliary + Subject personal pronoun + *not*) in the last two Middle English periods (1350–1500) in the *Helsinki Corpus* and 18 examples in the first Early Modern English period (1500–1570). He argues that the early appearance of this order and its increasing frequency over time is due to the weakening and cliticization of *not*, not to the decliticization of *thou*. For van Kemenade (2000: 70–71), too, the appearance of this order, which occurs, significantly, only after the loss of *ne*, is evidence for the incorporation of *not* as a clitic on the finite verb head.

3.4 (Early) Modern English

By 1500, *-tou/-tow* clitics seem to disappear. None of my Early Modern English sources (HC, CHVD, U of V, plus the *OED* quotations) provide any examples of enclitic forms, while at the same time they provide numerous examples of non-encliticized V + *thou* sequences. Thus, this would seem to be a clear case of decliticization, as Kroch et al. (1982) argue.

However, the process of subject enclisis itself is not lost. Newly formed clitics of the form *-ee*, *'ee*, *-y(e)*, derived from the 2nd person plural pronoun (*ye, you*), arise in this period (see *OED*: s.v.v. *ye*, def. A.1b, and *'ee*). Among closed-class items, this enclitic is found occasionally with BE and WILL in the 17th century (see (16)). Moreover, it occurs with great frequency with DO from the 17th to the early 20th century, especially in the representation of colloquial speech (*d'ee*, *d'ye*, *de'e*; *OED*: s.v. *d'ee*). Of the over 850 instances that I have found, (17) gives the variant forms.[14]

(16) a. Christalla, Philema where **are'ee**? (CHVD; 1633 Ford, *The Broken Heart* III, v, 107)[15]
b. Whither **will'ee**, / Without respect of shame? (CHVD; 1634 Ford, *The Chronicle Historie of Perkin Warbeck* V, iii, 80–81)
c. What **would'ee**? (CHVD; 1633 Ford, *'Tis Pitty Shee's a Whore* III, ix, 24)

(17) a. Here, here! Why **d'ye** baul so, father? **D'ye** think we have no ears? (HC; 1707 Farquahar, *The Beaux' Stratagem* I, i, 3–4)
b. Yes; how **d'ee** sister? …
D'ee mock mee, or flatter mee (CHVD; 1633 Ford, *'Tis Pitty Shee's a Whore* I, ii, 162; I, ii, 197)
c. Oh yes, and here agen, **de'e** not see me, you are so light / your selfe (CHVD; 1629 Shirley, *The Wedding* II, i, 17–18)

The host of the enclitic is typically auxiliary *do*, though note the set expression "how d'ee" in (17b), in which the host is main verb *do*.

What is more interesting, perhaps, is the occurrence of 2nd person enclitics with a new class of verbs, namely, verbs of attention (*hark, look, mind*) in the imperative.[16] *Hark'ee* (*hark'ye*, *harkee*; see *OED*: s.v. *hark*, def. 2c) is quite common in the 18th and 19th centuries. The form is also found in both British and American regional dialects (*DARE*: s.v. *hark*, def. 1b; *EDD*: s.v. *hark*, def. 1). It is used parenthetically, often before a vocative, as a discourse marker, a kind of attention-getter (without its original aural meaning).

(18) a. **Hark'ee**, friend, dost thou know this lady? (CHVD; 1773 Garrick, *The Chances* IV, iii, 48)
b. **Harkee**, sweet Mistress, how long have you practis'd / This Subtle Trade? (CHVD; 1700 Centlivre, *The Perjur'd Husband* I, ii, 200–201)
c. Here, **harkye**, sirrah! / Deny it all again (CHVD; 1765 Colman, *The Eunuch* IV, iv, 88)

The occurrence of a second *you* in some examples seems to suggest that the encliticized form is no longer analyzed as such and has become lexicalized.

(19) a. Ho! Here's Lucy coming — **Harkee you**, pray, why did you make me wait so long? (CHVD; 1700 Centlivre, *The Perjur'd Husband* IV, ii, 4–5)
b. **Harkye, you** mister Jack Joker! how came you in the midst of the prisoners! (U of V; 1823 Cooper, *The Pilot* Ch. 30)

Look'ee (*lookee, look-ee, looky, lookye, look'ye*; *OED*: s.v. *look*, def. 4a; *DARE*: s.v. *look*, def. 1) occurs in the same context as *hark'ee* with a similar discourse-marking function.[17]

(20) a. **Look'ye**, Doctor, you come before your time; I'ant condemn'd yet, I thank'ye (HC; 1707 Farquahar, *The Beaux' Stratagem* V, ii, 158–159)
b. **Lookye**, old gentleman. / If you've a liking to them, you must down / With the hard money (CHVD; 1767 Thornton, *The Shipwreck* III, iv, 40–42)
c. **Look'ee**, Jack, I have heard thee sometimes talk like an Oracle (*OED*; 1710 Steele, *Tatler* No. 206, p. 2)
d. **Lookee**, Frederick, / You talked to me of wenching. Let's have fair play (CHVD; 1773 Garrick, *The Chances* II, ii, 172–173)
e. And so they shall — **Looky'**, Lord Bishop (CHVD; 1746 Macklin, *King Henry VII* IV, iii, 29)

Looky is also found in regional dialects of both American and British English (*DARE*: s.v. *look*, def. A1; *EDD*: s.v. *look,* def. 12). An interesting development in the 19th century is the further reduced form, *look-a-here, lookyhere*, which appears in American sources.

(21) a. And now **look-a-here**; you ain't got now manner of business foolin' around here just now (U of V; 1897 Crane, "A Man and Some Others", *Century Monthly Magazine* 53 [1897]:604)
b. **Lookyhere**, Huck, what fools we're making of ourselves (U of V; 1876 Twain, *The Adventures of Tom Sawyer* Ch. 33)

Mind (you) is also used in the imperative "to call attention to, or emphasize, what the speaker is saying" (*OED*: s.v. *mind*, def. 5c). Though *mind you* is very common in Modern English, I have found only two pre-nineteenth century examples and only one (late) instance of encliticized *mind'ee*:

(22) a. **Mind you**, — when the captain / Comes home, be sure remember not to call / *Philocomastiium* by her name. (CHVD; 1767 B. Thornton, *The Braggard Captain* III, iii, 3–5)
b. But, no raptures / **mind you**. The walls are so unus'd to any sound / of joy, it might have a very sad effect upon them (CHVD; 1797 Boaden, *The Italian Monk* III, 250–252)
c. Be quickish downstairs, **mind'ee** (*OED*; 1900 Angus, *Jan Oxger* iii, 194)

3.5 Present-day English

In Present-day English, encliticized 2nd person subject forms continue to exist. Forms such as *would'ya, could'ya, did'ya* are common in colloquial speech. However, because of standard spelling practices, they are difficult to attest; for example, the British National Corpus contains only 6 examples of *would'ya*, 3 of *could'ya*, and 27 of *did'ya.*

4. Other clitics

A number of other types of 2nd person clitics are also attested. There are very rare examples of 2nd person **proclitics** in Middle English. I have found a few examples of *þart/þert* and *th'art* (< *thou art*) (see (23a–b)) as well as one example of *thall* (< *thou will*) (see (23c)).

(23) a. þer wot no man of wham **þart** come (*MED*; Mannyng, *Chronicle Pt.1* [Lamb 131] 8015)
'there knows no man of whom you are come'
b. **Th'art** wis ynough; forthi do nat amys (Chaucer, *TC* 3.1629)
'You are wise enough; therefore, do not do amiss'
c. Bot for all þat euer þou kane / **Thall** neuer be thryfty man (*MED*; c1500 *The Debate of the Carpenter's Tools* [Ashm 61] 5–6)
'but for all that ever you can, you will never be a thrifty man'

In Early Modern English, subject proclitics are common for a few forms,

especially *y'are* (*OED*: s.v. *y'are, yare*), *y'have,* and *y'had*, in which the host may be either an auxiliary or a main verb (see (24a–f)), while *y'know* and *y'were* are rare (see (24g–h)).

(24) a. **Y'are** practiz'd vpon most deuillishly (CHVD; 1623 Webster, *The Deuils Law-case* IV, ii, 289)
b. Egad, Sir, I think **y'are** in the right on't (HC; 1696 Vanburgh, *The Relapse* III, iii, 24–25)
c. In the rich Soyl, the sprightly Horse **y'have** seen, / Run, leap, and wanton o're the flow'ry green / Praunce, and curvet, with pleasure to the sight" (*OED*; 1682 Shadwell, *The Medal of John Bayes* 4)
d. **Y'have** a very fine brandenburgh on, Sir Fopling" (*OED*; 1676 Etheredge, *Man of Mode* IV, ii, 104)
e. I could have onely wishd you had acquainted / Her father, whom it equally concerns, / Though **y'had** presumed on me (CHVD; 1647 Fletcher, *Love's Pilgrimage* V, vi, 4–6)
f. Wherein I might both best and properliest / Discover my abusers, and your own, / And show you some content, before **y'had** none (CHVD; 1657 Middleton, *No Wit, No Help Like a Woman's* IV, iii, 159–161)
g. They've had no time / For a discovery yet; do **y'know** these, Madam? (CHVD; 1664 Stapylton, *The Step-Mother* V, 398–399)
h. He saies **y'were** moulded out of the first Earth (CHVD; 1651 Cartwright, *The Siege* IV, i, 58)

A different phenomenon is the occurrence of 2nd person **object** enclitics. The best known example is *prithee* (*prethe, preythe, prythee, prythy*) from *pray* + 2nd person singular *thee*. This form is common in the 17th and 18th centuries (*OED*: s.v. *prithee* — the earliest *OED* example is 1577, the latest 1875). In the Shakespearean corpus, Busse (1999: 86) has found 228 examples of *prithee* compared to 92 examples of *pray* followed by a non-enclitic 2nd person singular pronoun.[18] Furthermore, in Shakespeare, there is about an equal split between *prithee* forms with the subject present (see (25a–b)) and those with the subject omitted (see (25c)).[19]

(25) a. I will not meet with you tomorrow night. / I **prithee**, Diomed, visit me no more (Busse 1999; 1609 Shakespeare, *Troilus and Cressida* V, ii, 75–76)
b. I **prethe** looke what market she hath made (HC; 1630 Middleton, *A Chaste Maid in Cheapside* II, ii, 177)

c. **Prithee**, I do not ask thee what thou did'st not, but what thou did'st (HC; 1685 *The Trial of Lady Alice Lisle* IV, 121C2)[20]

For Busse (1999: 91), the occurrence of *prithee* in the same context as *you* (see (25a)) is evidence that *prithee* has become a "new monomorphemic unit" with cliticized *thee*. Furthermore, *prithee* occurs almost exclusively as a parenthetical (see (25c)) and functions as a discourse marker introducing indirect requests and questions (Busse 1999: 491–493; Traugott and Dasher 2002: 254–255). *Prithee* is seen as the final step in the grammaticalization of the verb *pray* (Busse 1999; Akimoto 2000).

A somewhat later host for the object enclitic is *thank* (*thankye, thank'ee, thankee, thanky*), from *thank* + the 2nd person plural *ye* (*OED*: s.v. *thankee*). Most examples date from the 19th century.

(26) a. I **thankee**, sir. I'm obleeged to you, sir, for your welcoming manner of me (U of V; 1850 Dickens, *David Copperfield* Ch. 7)
b. "**Thank'ee**, Jukes, **thank'ee**", would mutter Captain MacWhirr, without looking up (U of V; 1902 Conrad, *Typhoon and Other Stories* Ch. 1, p. 4)
c. **Thankye**; take it yourself — it will do you good (CHVD; 1903 Gilbert, *The Mountebanks* I, 802)
d. Mister Coxswain! **thanky**, Sir for giving me a handle to my name (*OED*; 1833 Marryat, *Peter Simple* Ch. IV)

(Also see (20a).) Again, the subject may or may not be present, and the expression is typically parenthetical.

According to Kroch et al. (1982: 290), the replacement of *ye*, "an extremely weak form", by the pre-existing strong form *you*, with eventual loss of *ye,* is further evidence of the decliticization of subject pronouns. However, they seem to be overlooking the cliticization of *ye* found in *d'ye*, *thank'ye*, and other forms in Early Modern English as well as the continued existence of *ye* well into the 18th century (see Lass 1999: 153–154). Moreover, the loss of *ye* would seem to be a complex change resulting from a number of external and internal factors (see Lutz 1998).

5. Accounting for *-tou/-tow* forms

The diachronic evidence just presented thus suggests no universal loss or demise of cliticization of 2nd person subject forms, but rather a continued

viability of subject enclisis and proclisis, as well of object enclisis, in the history of English.

However, the existence of *-tou/-tow* forms and their loss and replacement by the strong *thou* forms in Early Modern English still raises questions:

– what is the status of these forms?
– why are they lost? and
– how does their loss relate to larger grammatical developments occurring in Early Modern English?

5.1 Status of the forms

This paper began with the assumption that the forms under consideration are clitics, but are they? The question of what defines a clitic is a vexed one, well beyond the scope of this paper.[21] However, a useful set of criteria — morphological, syntactic, semantic, phonological, and lexical — is proposed by Sadock (1991: 52). The forms, *-tou/-tow*, *'ee*, and *'ya*, would appear to meet most of his criteria. Morphologically, they are bound elements which attach outside inflections and block further morphology to the right. Syntactically, they occur adjacent to their host[22] and are independent elements of syntax.[23] Phonologically, they are dependent, agglutinative, and stressless forms. Semantically, they serve a (grammatical?) function and take the meaning of a phrase as their argument. Lexically, they alternate with free words[24] and are not lexicalized (by which Sadock means unpredictable in meaning). In a couple of respects, however, they do not meet Sadock's criteria: they do not attach without regard to the morphological class of the host (though Sadock admits that many clitics violate this criterion [1991: 58]), and they are not completely productive. Sadock points out, though, that many "otherwise clear clitics" can differ in respect to almost any of the properties he lists (1991: 54).

The phonological derivation of the forms under consideration here may shed further light on their clitic or non-clitic status, though there seems to be some disagreement among the experts. First, Sadock argues that clitics are "subject to automatic phonological rules only" (1991: 52), but Nespor (1999: 875–880), in a discussion of "phonological clitics", argues that clitics are **not** derivable by regular phonological rules, though reduced forms are.[25] Citing the example of the English auxiliary *'s*, she notes that while vowel reduction is an independently motivated rule of English ([ɪz] > [əz]), "there is no rule that, say deletes a vowel before [z]" (875). Second, in distinguishing clitics from independent words, Zwicky (1985: 286) argues that clitics must be subject to

internal, not external sandhi rules.[26] However, Nespor (1999: 868–872) shows that internal rules need not apply to clitics; rather, clitics constitute a special phonological domain. In our case, the sandhi rule assimilating thorn to the preceding dental, as in *þou* > *tou*, is a regular phonological process that is both word internal and external (see Jordan 1974: 186).[27] Likewise, palatalization of [d] > [dʒ], as in *would'ya*, is a regular phonological process that is both internal and external. However, the development of [ji/jə] to [i] or [ə], as in *lookee*, would not appear to be a regular or independently motivated phonological process in English (though [ji] > [jə] is entirely natural).

The phonological evidence suggests, therefore, that *-tou/-tow* and *'ya* are not true clitics but merely reduced or contracted forms. Only *'ee* might constitute a true clitic. Some evidence for this is the occurrence of *'ee* with a full range of verbs and, more importantly, with parts of speech other than verbs, such as *t'ee* < *to* + *ye* (see (27a)), *w'ee* < *with* + *ye* (see (27b)), or *int'ee* < *into* + *ye* (see (27c)). The first two appear to be relatively common.

(27) a. you forsooth / Must haue your loue come **t'ee** (CHVD; 1606 Chapman, *The Gentleman Usher* V, i, 19–20)
b. You must excuse me friend I would Joyn **w'ee** (CHVD; 1653 Shirley, *The Brothers* IV, ii, 16)
c. The hundred thousand knacks not to be spoken of / (And all this for King Oberon and Queene Mab) / Should put a soule **int'ee** (CHVD; 1634 Ford, *The Chronicle Historie of Perkin Warbeck* III, ii, 11–13)

The occurrence of *harkee* in the presence of *you* (see (19)) might be further evidence of the clitic status of *'ee*. Note that the *lookyhere/lookahere* forms do not violate criteria for clitichood, since clitics, though they prevent the addition of further morphology, may be followed by other clitics.

5.2 Loss of the forms

Putting aside the status of *-tou/-tow* forms for the moment, let's consider what process might account for their loss. Is it decliticization, as far as we understand the process? If we follow Lehmann's stipulation (see above, Section 2) that decliticization must involve forms which are exclusively clitic, then we are clearly not dealing with a case of decliticization since these clitic (or reduced) forms have always been optional variants of the corresponding full forms. Even Jeffers and Zwicky's weaker claim (see above, Section 2) that decliticization involves the emergence or re-emergence of an independent word from a clitic

would seem to imply a period when the independent form ceased to exist. This is not the case here.

In fact, there is an overarching problem with viewing decliticization as the mirror-image or reverse of cliticization. Cliticization typically involves not only morphological bonding but also some degree of phonological attrition, and as Janda (2001:295) notes, phonological strengthening processes, such as the change from reduced vowel (schwa) to full vowel, are very rare, if not non-existent. Norde (2001:236) echoes Janda in pointing out that the "mirror-image" process to cliticization would be "logically impossible, since grammaticalization frequently involves semantic and phonological reduction", and we are unable to predict full forms from reduced forms. The change from *thou* to *-tou*, though not a case of vowel reduction, involves assimilatory changes that would be unlikely to be reversed, nor could the original form be recovered. Norde, therefore, opts instead to consider degrammaticalization as a "counter-directional" movement, from right to left (more to less grammatical) on the cline of grammaticality (2001:237). Janda too chooses to understand degrammaticalization as a case where grammaticalization has been "countered (or counteracted) in such [a] way as to be canceled out — i.e. returned to an earlier starting point — by one or more processes" (2001:296). While both of these suggestions allow something to be said about the starting point and the end point of degrammaticalization, however, they still do not provide insight into the process (the "how") or the motivation (the "why") for the change.

A clue to the change we are witnessing here is provided by Bybee, Perkins, and Pagliuca (1994:13), who point out that "once phonological segments are reduced or deleted from gram[matical element]s, the grams do not again assume their fuller form *unless that fuller form has also been preserved in the language and replaces the reduced form*" (my emphases). That is, the reappearance of full *thou* forms in 16th century English after a period of reduced *-tou/-tow* forms is a instance not of "decliticization", but of "replacement" by a pre-existing form, one that has continued to exist. Like replacement processes in general, this is a sudden change with no intermediate states — as opposed to degrammaticalization processes, which are usually seen as gradual or step-wise involving a complex of interrelated semantic, morphological, syntactic, and phonological changes. Furthermore, I would suggest that this replacement is motivated, or necessitated, by various phonological changes in Early Modern English, specifically, loss of the sandhi rule of assimilation of thorn to the preceding dental. Evidence that this rule ceases to operate is the absence of examples of assimilated demonstratives, articles, or other words beginning in

thorn in this period. Concomitantly, there is the emergence of the palatalization rule of /d/ > /dʒ/ before a weak syllable containing /i, j/ which arose in the 15th century but became fully established mainly in the 17th c. (see Lass 1999: 120). This produced the frequent *d'ye* forms (from *do you*) of Early Modern English as well as the later *'ya* forms, as in *would'ya* or *could'ya*, which coexist with but do not replace the full forms in colloquial speech.[28]

5.3 Relation to larger grammatical developments

Kroch et al. (1982) see a direct correlation between the (a) cliticization of *thou* and the low rate of *do* in questions in the early 16th century, and (b) decliticization of subject pronouns and the spread of *do* to questions with pronominal subjects in the later part of the 16th century.

In respect to (a), it is possible that questions with 2nd person singular subjects — where enclisis is common — might constitute the questions without *do* in Kroch et al.'s (1982) data, while questions with 1st and 3rd person, as well as 2nd person plural, subjects, where enclisis in uncommon, might constitute those with *do*. However, data in Stein (1990: 168–169, 184–187) for both the Shakespeare corpus and for earlier and later 16th–17th century corpora do not seem to corroborate such an assumption. First, questions with 2nd person singular subjects are much less common than questions with 2nd person plural subjects or non-2nd person subjects (approximately 1/3 less common than either in Shakespeare). Second, rather than showing the lowest rate of *do*-periphrasis, questions with 2nd person subjects show the highest rate in Stein's data.

In respect to (b), as I have argued above, the loss of *-tou/-tow* forms can probably be motivated by loss of a phonological rule. Nonetheless, there are potentially three larger grammatical changes which might interact with loss of these subject clitics: (1) the replacement of *thou* by *you* as the unmarked 2nd person pronoun; (2) the non-expression of subjects in imperatives, and (3) the spread of *do* in questions. All of these changes are complex in nature, and it is well beyond the scope of this paper to treat them.[29] However, suffice it to say that the loss of the *tou/-tow* forms seems to have predated and to have been well completed before any of these changes. The latest examples that I have found of subject *-tou/-tow* forms date from 1500, with most occurring before 1450, whereas these larger grammatical changes do not seem to have been completed until the 17th century, or later. Furthermore, while the grammatical changes seem to have been gradual changes extending over several centuries, the loss of *-tou/-tow* forms seems to have been much more sudden and cataclysmic.

6. Conclusion

In conclusion, I believe that this study has shown that *-tou/-tow* in forms such as *hastou* and *wiltou* are optional reduced forms, not clitics. Their loss in 16th century English comes about through a process of replacement (by a pre-existent full form), not through decliticization, and is an instantaneous rather than a gradual change. Furthermore, replacement is motivated by loss of the sandhi rule assimilating thorn to a preceding dental and is not related to larger grammatical changes in the language. Returning to Newmeyer's (1998) use of this example, therefore, it seems clear that it will not serve as a valid counter-example to the tenet of unidirectionality in grammaticalization, as these forms are neither clitics nor decliticized. Finally, I believe that it casts doubt on the process of decliticization as a legitimate type of degrammaticalization, but this is a topic requiring further research.

Notes

* I would like to acknowledge UBC-SHRCC for providing a travel grant which enabled me to attend the 15th International Conference on Historical Linguistics in Melbourne in August 2001. I am grateful to members of the audience, including Theo Vennemann, Ans van Kemenade, Donka Minkova, Brian Joseph, and Marie-Lucie Tarpent, who allowed me to focus my ideas and make significant revisions to the paper. I am also grateful to the Medieval English Society of Japan for supporting my trip to Japan in March 2002 and to the audience at Seikei University, Tokyo, for stimulating discussion.

1. The second half of 'demorphologization', namely, the "phonologization of a once morphologically determined alternation" is **not** decliticization, but rather what Hopper and Traugott (1993: 164–164) term "phonogenesis". Two of the examples cited by Joseph and Janda (1988: 198) fall into this category.

2. I was unable to obtain Rubino (1994), but see Campbell (2001: 128) and Newmeyer (1998: 271–272) for discussions of this article.

3. Jeffers and Zwicky (1980: 224) seem to attribute decliticization to a kind of *drift*, or a "shift away from the use of clitics toward a greater dependence on isolated words".

4. According to Janda (1980), the group genitive clitic does not develop directly from the inflection, but rather via the *his* genitive, a reanalyzed form of the inflection which he interprets as an enclitic; the spread of the group genitive exactly parallels that of the *his* genitive. He sees the demise of the *his* genitive as due to the increasing non-syllabicity of the form as well as stylists' attacks on it. Allen (1997) contests Janda's view, citing an impressive array of data. She argues that the *his* genitive (what she calls the "separated genitive") is merely an orthographic variant of the inflected ("attached") genitive, sharing the same

distribution; the group genitive is a direct development of the inflected genitive in its spread to all noun classes. Group *his*-genitives do not appear until after group inflected genitives are well established. She notes that only later are there attempts to treat the form in the *his*-genitive as a pronoun (hence, the appearance of *her* and *their* in this structure). Allen notes (2003) that it may not be possible to make a sharp distinction between inflection and clitic in this case. See also Lehmann (1995: 18–19).

5. Some 1st person examples are as follows:

a. Giftes **willi** geue þe / Þat þou maiȝt euer betere be (HC; *Dame Sirith* 388–389)
'I will give you gifts so that you might ever beter be'

b. **Herdi** neuere bi no leuedi / Bote hendinese and curteysi / And ioye hy gunnen me bringe (HC; *The Thrush and the Nightengale* 101–103)
'I never heard by no lady but gentleness and courtesy / And joy they brought me'

c. **Cani** do non oyir dede, / Bot my pater noster and my crede (HC; *Interlude* [Appendix to *Dame Sirith*] 71–72)
'I can do no other deed except my pater noster and my credo'

d. ffor aly wyman **ami** on (HC; *Interlude* [Appendix to *Dame Sirith*] 85)
'for all women am I one'

The paucity of first-person clitics can be explained pragmatically by the oddness of questions directed towards oneself, but a similar explanation does not explain the paucity of third-person clitics.

6. Strang's explanation for their disappearance — that they are lost because of "a growing sense that writing should be ideographic, i.e., that a single form should stand for each word in all circumstances" (1970: 263) — seems rather odd.

7. No precise comparison is intended here as the corpora differ in size: the Middle English section of the *Helsinki Corpus* contains 608,570 words; Chaucer's complete works contain 428,915 words, and the Middle English section of the University of Virginia Collection contains 65 titles, "most of which are publicly accessible". However, given that the Chaucerian corpus is approximately 30% smaller than the *Helsinki Corpus*, one can see the popularity of enclitic forms in Chaucer's language.

8. These figures can only be interpreted as approximate because in the case of frequent forms, multiple uses of the same quotation make it very difficult to arrive at an exact count.

9. All Chaucerian quotations and title abbreviations follow Benson (1987). To the extent that it was possible, examples were checked against printed versions. The Wycliffe Bible quotations were checked again the online "Bible in English".

10. Multiple instances of *canstow* in the quotations in the *MED* all come from the same passage in *Piers Plowman*.

11. *The English Dialect Dictionary* (Wright 1898–1904; henceforth *EDD*) cites the modern variants *dusta*, *dusto*.

12. As remnants of *see thou*, the *EDD* cites the forms *sithee/sithe and seesta*, an exclamation used to attract attention. It also records the exclamation *see-yo*, derived from *see you* (see below, on 2nd person plural enclitics).

13. Interestingly, Gower, a contemporary of Chaucer, writing in roughly the same dialect, consistently uses non-encliticized forms. Whether this is a matter of personal preference, the result of minor dialectal differences, an accident of scribal practice, or the consequence of modern editorial practice is difficult to ascertain; discussions of such issues in Chaucer and Gower (see, e.g., Smith 1988) make no mention of this feature.

It is perhaps also significant that most (but not all) of the non-encliticized forms in Chaucer occur in prose, suggesting a metrical motivation, at least in part, for enclisis. However, if this were the primary motivation, one would expect to find enclisis in Gower as well.

14. See also *EDD* (s.v. *do* [*d'ye, didst'ee*]). Cassidy and Hall (1996, *Dictionary of American Regional English,* henceforth *DARE*; s.v. *'ee*) sites the forms *hav'ee*, *do'ee*, *sit'ee*, *tell'ee*, *dost'ee*, *wi'ye*, and *hae ee*.

15. In regard to examples of *are'ee*, there is a hiatus in my data until the 19th century.

16. Use of a postposed pronoun subject with the imperative goes back to Old English, but in Modern English is now restricted to set phrases such as *mind you* or *believe you me* or to dialects (Visser 1963: 16–17). The appearance of the subject in Early Modern English is dependent on various syntactic factors; e.g., an unstressed object pronoun prevents the appearance of the pronominal subject (Rissanen 1999: 278). According to Denison (1998: 249), the major change in the modern period is disuse of this pattern and the rise of an alternative pattern with *you* before the verb (see also Visser 1963: 17–18).

While both *harkee* and *lookee* are traditionally said to derive from the verb and a cliticized plural nominative *ye*, constructions with *thee*, which are common in Early Modern English, could likewise be the source. Millward (1966) points out that verbs of attention (*hark*, *hear*, *look*, *mark*) are always accompanied by the objective *thee*, as are verbs of motion (*come*, *haste*, *hie*, *return*, *run*, etc.) and verbs taking a reflexive direct or indirect object (*get*, *make*, *seek*, *seize*, *take*, etc.). Abbott attempts to explain the appearance of the objective with verbs of attention as follows: "*Thee*, thus used, follows imperatives which, being themselves emphatic, require an unemphatic pronoun. The Elizabethans reduced *thou* to *thee*" (1966: 141).

17. For further discussion of forms of *look*, see Brinton (2001).

18. 393 examples of non-enclitic 2nd person plural pronouns (*you/ye*) following *pray* also occur in Shakespeare (Busse 1999: 486).

19. Jespersen ([1961]: 224) calls the process of subject omission found in forms such as *thank you* or *pray you* "prosiopesis".

20. See Kryk-Kastovsky (2000: 213–216) for a discussion of the function of *prithee* and other discourse markers in this trial as well as the trial of Titus Oates.

21. See van Riemsdijk (1999) for a useful overview of the state of clitic studies, especially in a generative context.

22. The question of whether the clitic occurs syntactically in the same position as the corresponding full form or in a different position distinguishes what Zwicky calls "simple clitics" form "special clitics" (1985: 295).

23. If we accept the view that all pronouns in Old English are (syntactic) clitics (van Kemenade 1987), the forms are not derived from syntactically independent elements. However, Koopman (1997) finds that though there is a gradient of clitic-like behavior among

Old English pronouns, which sometimes show different syntactic behavior from full NPs, no pronoun satisfies all of the criteria for clitichood (proposed by R.S. Kayne for French clitics). See also Fischer, van Kemenade, Koopman, and van der Wurff (2000: 199–120).

24. Many "special clitics" (see note 22), such as the object clitics in French, are obligatory.

25. Clitics and reduced forms may have different distributions, see Nespor (1999: 876).

26. See van Riemsdijk (1999: 7–12) for further discussion of Zwicky's criteria.

27. Word internally, this type of assimilation occurred (already in Old English) in syncopated 3rd person singular present forms such as *fint*, *get*, *bit*, and *stant*, as well as in compounds such as *West* + *þyriding* > *Westtriding* (Jordan 1974: 186).

28. The *harkee* and *lookee* forms, which arguably contain true clitics, have been lost though a process of lexical attrition, not decliticization.

29. Much scholarly attention has been focussed especially on the regularization of *do* (see Rissanen 1999b: 239–248 for an overview).

Appendix

(1) a. A, son, vat **saystu**? Benedicite! / Lift hup yi hand and blis ye! (HC; *Interlude* [Appendix to *Dame Sirith*] 64–65)
'Ah, son, what do you say? Bless you! Lift up your hand and bless yourself'

b. O suster deere, / Ne **seydestow** right now in this manere (Chaucer, *CT* G.SN 333–334)
'Oh sister dear, you did no say right now in this manner'

c. Certes thus **seestow** wel that manye thynges failen to hym (HC; Chaucer, *Bo* 3.pr.9.108–109)
'Certainly, thus you see well that many things fail to him'

d. That **knowestow** wel thyself, but if thou madde. / Why artow angry with my tale now? (Chaucer, *CT* A.Mil 3156–3157)
'You know that well yourself if you go mad. Why are you angry with my tale now?'

(2) a. But **axestow** in somme of what gylt I am accused? (Chaucer, *Bo* 1.pr.4.138–139)
'But are you asking in a few words of what guilt I am accused?'

b. O Maystresse, what **demestow** of this? (Chaucer, *Bo* 1.pr.4.145)
'Oh Mistress, what do you judge of this?'

c. And **desirestow** to heren in what manere? (Chacuer *Bo* 1.pr.4.140–141)
'And do you desire to hear in what way?'

d. What! Alison! **Herestow** nat Absolon? (Chaucer, *CT* A.Mil 3366)
'What! Alison! Don't you hear Absolon?'

e. **Holdestow** thanne thilke welefulnesse precious to the? (Chaucer, *Bo* 2.pr.1.73–74)
'Do you hold then this happiness precious to yourself?'

f. Whi **listow** in this wise, / Syn thi desir al holly hastow had? (Chaucer, *TC* 4.394–395)
'Why are you lying down in this way since you have had your desire completely?'

g. Thyn instrument so … / **Makestow** of wommen whan thou wolt bigile (Chaucer, *CT* B.ML 370–371)
'You make of women your instrument when you will deceive'
h. **Remembrestow** what is the ende of thynges? (Chaucer, *Bo* 1.pr.6.37–38)
'Do you remember what is the end of things?'
i. What **spekestow** of preambulacioun? (Chaucer, *CT* D.WB 837)
'Why do you speak of making a preamble?'
j. Ne **takestow** noon heede / Of my power? (Chaucer, *CT* G.SN 435–436)
'Don't you take any heed of my power?'
k. Quod Pandarus, "What **thynkestow** to do?" (Chaucer, *TC* 4.849)
'Said Pandarus, "What do you think to do?"'
l. For **trowestow** that Philosophie be now alderferst assailed in periles by folk of wykkide maneris? (Chaucer, *Bo* 1.pr.3.21–22)
'For do you believe that Philosophy is now first of all assailed in perils by people of a wicked kind'
m. This to seyn, "**Undirstondestow** aught that soules han any torment aftir the deeth of the body?" (Chaucer, *Bo*. 4.pr.4.145–146)
'This to say, "Do you understand at all that souls have any torment after the death of the body?"'
n. What, **wenestow** make an ydiot of oure dame? (Chaucer, *CT* D.WB 311)
'What, do you expect to make an idiot of our lady?'

(3) a. Mi Sone, **art thou** knowende of this? (*MED*; Gower, *Confessio Amantis* [Frf 3] 3.864)
'My son, do you know of this?'
b. for Latyn ne **canst thou** yit but small, my litel sone (*MED*; Chaucer, *Astr* pr. 27–28)
'for Latin, you do not know yet but a small amount, my little son'
c. Let see! **Darst thou** yet loke now? (*MED*; Chaucer, *HF* 580)
'Let it be seen! Do you dare yet to look now?'
d. In what power **dost thou** these thingis, and who ȝaf to thee this power? (*MED*; *WBible (1)* [Dc 369(2)] Matthew 21.23)
'In what power do you do these things, and who gave to you this power?'
e. Than **hast thou** a brod reule (*MED*; Chaucer, *Astr* 1.13.1)
'Then you have a broad aliolade'
f. And by thys conclusioun **maist thou** take ensample in alle the signes (*MED*; Chaucer, *Astr* 1.21.67–68)
'And from this conclusion you may take example in all the signs'
g. How **mihtest thou** thin herte finde / … That thou acordest to the slawhte / Of him which was thin oghne lord? (*MED*; Gower, *Confessio Amantis* [Frf 3] 3.2056–2059)
'How might you find your own heart so that you reconcile to the slaying of him who was your own lord?'
h. and of hem **shalt thou** aske Thy conseil (U of V; Chaucer, *CT* B.Mel 1155)
'and of them you shall ask your counsel'
i. **Wilt thou** lere of sterres aught? (*MED*; Chaucer, *HF* 993)
'Will you learn anything of stars?'

j. Thus **wost thou** wel what thing is speche (*MED*; Chaucer, *HF* 773)
'thus do you know well what a thing speech is'

(4) a. **Knowest thou** ouht A Corseynt Men calleþ Seynt Treuþe? (*MED*; *Piers Plowman.*A(1) [Vrn] 6.23)
'Do you know anything of a Corsaint men call Saint Truth?'

b. What **seest thou** in thi brotheris yȝe a festu [L festucam] othir a mot? (*MED*; *WBible*(1) [Dc 369(2)] Luke 6.41)
'What do you see in your brother's eye a mote or a speck?'

c. Where **trowest þou** he schal be cast þat haþ reued oþer mennes from hem? (*MED*; Wimbledon *Sermon* [Corp-C 357] 87/411–412)
'Where do you think he shall be cast who has deprived other men of them?'

d. How liȝtly **seist thou** thee a Romayn citeseyn? I with moche summe gat this ciuylite [*WB(2)*: fredom; L civilitatem]. (*MED*; *WBible(1)* [Dc 369(2)] Deeds 22.28)
'How lightly do you say yourself a Roman citizen? I with a great sum got this civility'

e. Whi **makest þou** swich scher and foul lour? (*MED*; *7 Sages(1)* [Auch] 86/1818)
'Why do you make such a mien and foul frowning countenance?'

f. **Wenst þou** þat ichulle fleoȝ / Nai par deu, nouȝt a fote: for þe king ne for þe! (*MED*; *The Early South-English Legendary,* "St. Thomas of Caunterbury" [LdMisc 108] 2045–2046)
'Do you think that I shall flee? Nay, par dieu, not a foot for the king nor for you!'

g. So brouke þou þi croune, Wi **comest þou** fram toune? (*MED*; *King Horn* [LdMisc 108] 1107–1108)
'So you broke your crown. Why do you come from town?'

h. Muche tale þou makest þe foule bodi to queme, / But of þe seli soule **Takest þow** no ȝeme. (*MED*; *A Disputison Bitwene a God Man and þe Devil* [Vrn] 786–787)
'A great fiction you make the foul body to please, but of the innocent soul you take no heed'

i. A louely ladi In linnene I-cloþed / Com a-doun from þe [clyf] and clepte me feire / And seide, "sone! **slepest þou**?" (*MED*; *Piers Plowman.A(1)* [Vrn] 1.4)
'A lovely lady clothed in linen came down from the cliff and called me fairly and said, "son, are you sleeping?"'

j. **Herdes þou** nevere the processe of þe actis of þe apostlis? (*MED*; *Friar Daw's Reply* [Dgb 41] 775)
'Never did you hear the process of the acts of the apostles?'

k. **Denyest þou** þat holy chirche schal knowe correctt men of synnes? (*MED*; Trevisa, *Dialogus inter Militem et Clericum* [Hrl 1900] 11/4)
'Do you deny that the holy church shall know men rid of sin?'

(5) a. On me **nast þou** power non: swych destresse for-to do (*MED*; *The Early South-English Legendary,* "St. Thomas of Caunterbury" [LdMisc 108] 761)
'On me you do not have any power such distress to do'

b. **Nylt þou** neuer to Nuniue bi no-kynnez wayez? (*MED*; *Patience* [Nero A.10]346)
'Will you never to Nunive by no means?'

References

Online Sources

Bible in English (http://www.library.ubc.ca/ssp/bie/)
British National Corpus (http://www.hcu.ox.ac.uk/BNC/)
Chadwyck Healey English Verse Drama (http: //www.library.ubc.ca/ssp/evd/)
Middle English Dictionary Online, Middle English Compendium (http: //ets.umdl.umich.edu/m/mec/)
Oxford English Dictionary. 3rd ed. online. (http://dictionary.oed.com/entrance.dtl)
University of Virginia Electronic Text Center, Middle and Modern English Collections. (http://etext.lib.virginia.edu/eng-on.html)

Print Sources

Abbott, E. A. 1870. *A Shakespearian Grammar*. (3rd revised ed.) New York: Macmillan and Company. [Reprinted by: Dover, 1966.]
Akimoto, M. 2000. "The grammaticalization of the verb 'pray'". In *Pathways of Change: Grammaticalization in English* (Studies in Language, Companion Series 53), O. Fischer, A. Rosenbach and D. Stein (eds.), 67–84. Amsterdam and Philadelphia: John Benjamins.
Allen, C. L. 1997. "The origins of the 'group genitive' in English". *Transactions of the Philological Society* (Special Number — Papers from the English Historical Syntax Meeting, Manchester 11–12 May 1996) 95.1: 111–131.
Allen, C. L. 2003. "Deflexion and the development of the genitive in English". *English Language and Linguistics* 7: 1–28.
Benson, L. D. (ed.). 1987. *The Riverside Chaucer*. (3rd ed.) Boston: Houghton Mifflin.
Brinton, L. J. 2001. "From matrix clause to pragmatic marker: The history of *look*-forms", *Journal of Historical Pragmatics* 2: 177–199.
Busse, U. 1999. "'Prithee now, say you will, and go about it.' *Prithee* vs. *pray you* as discourse markers in the Shakespeare corpus". In *Anglistentag 1998: Erfurt, Proceedings*, F.-W. Neumann and S. Schülting (eds.), 485–500. Trier: WVT Wissenschaftlicher Verlag Trier.
Bybee, J., Perkins, R. and Pagliuca, W. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago and London: University of Chicago Press.
Campbell, L. 1991. "Some grammaticalization changes in Estonian and their implications". In *Approaches to Grammaticalization, Vol. 1: Focus on Theoretical and Methodological Issues* (Typological Studies in Language 19: 1), E. C. Traugott and B. Heine (eds.), 285–299. Amsterdam and Philadelphia: John Benjamins.
Campbell, L. 2001. "What's wrong with grammaticalization?". *Language Sciences* 23: 113–161.
Cassidy, F. G. and Hall, J. H. (eds.). 1996. *Dictionary of American Regional English, Vol. III*. Cambridge, MA and London: Belknap Press. (= *DARE*)
Denison, D. 1998. "Syntax". In *The Cambridge History of the English Language, Vol. IV: 1776–1997*, S. Romaine (ed.), 92–329. Cambridge: Cambridge University Press.

Fischer, O., van Kemenade, A., Koopman, W. and van der Wurff, W. 2000. *The Syntax of Early English.* (Cambridge Syntax Guides.) Cambridge: Cambridge University Press.

Hogg, R.M. 1992. "Phonology and morphology". In *The Cambridge History of the English Language, Vol. I: The Beginnings to 1066*, R.M. Hogg (ed.), 67–167. Cambridge: Cambridge University Press.

Hopper, P.J. and Traugott, E.C. 1993. *Grammaticalization.* (Cambridge Textbooks in Linguistics.) Cambridge: Cambridge University Press.

Janda, R.D. 1980. "On the decline of declensional systems: The overall loss of OE nominal case inflections and the ME reanalysis of *-es* as *his*". In E.C. Traugott et al. (eds.), 243–252.

Janda, R.D. 2001. "Beyond 'pathways' and 'unidirectionality': On the discontinuity of language transmission and the counterability of grammaticalization". *Language Sciences* 23: 265–340.

Jeffers, R.J. and Zwicky, A.M. 1980. "The evolution of clitics". In E.C. Traugott et al. (eds.), 221–231.

Jespersen, O. [1961]. *A Modern English Grammar on Historical Principles, Part 3: Syntax (second volume).* London: George Allen and Unwin/Copenhagen: Ejnar Munksgaard.

Jordan, R. 1974 [1934/1925]. *Handbook of Middle English Grammar: Phonology,* E.J. Crook (trans. and rev.). The Hague and Paris: Mouton.

Joseph, B.D. (ed.). 1986. *Studies in Language Change.* (Working Papers in Linguistics 34.) Columbus: Department of Linguistics, The Ohio State University.

Joseph, B.D. and Janda, R.D. 1988. "The how and why of diachronic morphologization and demorphologization". In *Theoretical Morphology: Approaches in Modern Linguistics*, M. Hammond and M. Noonan (eds.), 193–210. San Diego: Academic Press.

Kemenade, A. van. 1987. *Syntactic Case and Morphological Case in the History of English.* Dordrecht, Holland and Providence, RI: Foris.

Kemenade, A. van. 2000. "Jespersen's Cycle revisited: Formal properties of grammaticalization". In *Diachronic Syntax: Models and Mechanisms*, S. Pintzuk, G. Tsoulas and A. Warner (eds.), 51–74. Oxford: Oxford University Press.

Koopman, W. 1990. *Word Order in Old English, with Special Reference to the Verb Phrase.* (Amsterdam Studies in Generative Grammar 1.) Amsterdam.

Koopman, W. 1997. "Another look at clitics in Old English". *Transactions of the Philological Society* (Special Number — Papers from the English Historical Syntax Meeting, Manchester 11–12 May 1996) 95.1: 73–93.

Kroch, A., Myhill, J. and Pintzuk, S. 1982. "Understanding *do*". In *Papers from the 18th Regional Meeting of the Chicago Linguistic Society*, K. Tuite et al. (eds.), 282–294. Chicago: Chicago Linguistic Society.

Kryk-Kastovsky, B. 2000. "Representations of orality in Early Modern English trial records". *Journal of Historical Pragmatics* 1: 201–230.

Kytö, M. 1996. *Manual to the Diachronic Part of the Helsinki Corpus of English Texts: Coding Conventions and Lists of Source Texts.* (3rd ed.) Helsinki: Department of English, University of Helsinki.

Lass, R. 1999. "Phonology and morphology". In R. Lass (ed.), 56–186.

Lass, R. (ed.). 1999. *The Cambridge History of the English Language, Vol. III: 1476–1776.* Cambridge: Cambridge University Press.

Lehmann, Ch. 1995. *Thoughts on Grammaticalization* (LINCOM Studies in Theoretical Linguistics 01.). München and Newcastle: Lincom Europa.

Lessau, D.A. 1994. *A Dictionary of Grammaticalization, Band 1.* (Bochum-Essener Beiträge zur Sprachwandelforschung, Band 21.) Bochum: Universitätsverlag Dr. N. Bockmeyer.

Lutz, A. 1998. "The interplay of external and internal factors in morphological restructuring: The case of *you*". In *Advances in English Historical Linguistics (1996)*, J. Fisiak and M. Krygier (eds.), 189–210. Berlin and New York: Mouton de Gruyter.

Matsumoto, Y. 1988. "From bound grammatical markers to free discourse markers: History of some Japanese connectives". In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistic Society*, S. Axmaker, A. Jaisser and H. Singmaster (eds.), 340–351. Berkeley: Berkeley Linguistics Society.

Millward, C. 1966. "Pronominal case in Shakespearean imperatives". *Language* 42: 10–17. [Reprinted in: V. Salmon and E. Burness (eds.), *A Reader in the Language of Shakespearean Drama.* Amsterdam and Philadelphia: John Benjamins, 1987, 301–308.]

Mossé, F. 1945. *Manuel de l'anglais du moyen âge: des orignes au XIVe siecle. I Vieil-Anglais.* (Bibliothèque de Philologie Germanique 8.) Paris: Editions Montaigne.

Mossé, F. 1952. *A Handbook of Middle English*, J.A. Walker (trans.). Baltimore and London: Johns Hopkins University Press.

Nespor, M. 1999. "The phonology of clitic groups". In: H. van Riemsdijk (ed.), 865–887.

Nevis, J.A. 1986a. "Decliticization and deaffixation in Saame: Abessive *taga*". In B. Joseph (ed.), 1–9.

Nevis, J.A. 1986b. "Decliticization in Old Estonian". In B. Joseph (ed.), 10–27.

Newmeyer, F.J. 1998. *Language Form and Language Function.* Cambridge, MA and London: The MIT Press.

Norde, M. 2001. "Deflexion as a counterdirectional factor in grammatical change". *Language Sciences* 23: 231–264.

Oizumi, A. 1991–92. *A Complete Concordance to the Works of Geoffrey Chaucer.* (Alpha-Omega Reihe C, Englische Autoren.) K. Miki (programmer). 10 vols. Hildesheim and New York: Olms-Weidmann.

Rissanen, M. 1999a. "*Isn't it?* or *is it not?* On the order of postverbal subject and negative particle in the history of English". In *Negation in the History of English* (Topics in English Linguistics 26), I. Tieken-Boon van Ostade, G. Tottie and W. van der Wurff (eds.), 189–205. Berlin and New York: Mouton de Gruyter.

Rissanen, M. 1999b. "Syntax". In R. Lass (ed.), 187–331.

Rubino, C. 1994. "Against the notion of unidirectionality in lexeme genesis". *Linguistica Atlantica* 16: 135–147.

Sadock, J.M. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Representations.* (Studies in Contemporary Linguistics.) Chicago and London: University of Chicago Press.

Smith, J.J. (ed.). 1988. *The English of Chaucer and his Contemporaries: Essays by M. L Samuels and J. J. Smith.* Aberdeen: Aberdeen University Press.

Stein, D. 1990. *The Semantics of Syntactic Change: Aspects of the Evolution of* do *in English.* (Trends in Linguistics, Studies and Monographs 47.) Berlin and New York: Mouton de Gruyter.

Strang, B. 1970. *A History of English.* London: Methuen.

Trask, R.L. 1993. *A Dictionary of Grammatical Terms in Linguistics.* London: Routledge.

Traugott, E. C. and Dasher, R. B. 2002. *Regularity in Semantic Change.* (Cambridge Studies in Linguistics 96.) Cambridge: Cambridge University Press.

Traugott, E. C., Labrum, R. and Shepherd, S. (eds.). 1980. *Papers from the 4th International Conference on Historical Linguistics.* (Current Issues in Linguistic Theory 14.) Amsterdam and Philadelphia: John Benjamins.

van Riemsdijk, H. 1999. "Clitics: A state-of-the-art report". In H. van Riemsdijk (ed.), 1–30.

van Riemsdijk, H. (ed.). 1999. *Clitics in the Languages of Europe.* (Empirical Approaches to Language Typology EUROTYP 20–5.) Berlin and New York: Mouton de Gruyter.

Venezky, R. L. and diPaolo Healey, A. (eds.). 1980. *A Microfiche Concordance to Old English.* (Publications of the Dictionary of Old English 1.) Newark, DL and Toronto: University of Delaware Press.

Visser, F. Th. 1963. *An Historical Syntax of the English Language, Part 1: Syntactical Units With One Verb.* Leiden: E. J. Brill.

Wright, J. 1898–1904. *The English Dialect Dictionary.* London: Henry Frowde (for the English Dialect Society) and New York: G. P. Putnams. (= *EDD*)

Zwicky, A. M. 1985. "Clitics and particles". *Language* 61: 283–305.

Name index

A
Aarts, B. 9
Abbott, E. A. 249 n16
Aijmer, K. xi, 77, 194
Akimoto, M. 242
Allen, C. L. 247–8 n4
Altenberg, B. 58, 74
Andersen, G. 10
Arce-Arenales, M. 114 n7
Aston, G. 174, 206 n2

B
Baayen, H. 17
Bäcklund, U. 23–4
Bailey, R. W. 101, 108, 110, 115, 137
Bailey, G. 214, 217, 219–20
Ball, C. N. 207 n13
Bassett. M 219, 220
Benson, L. D. 248 n9
Berglund, Y. 10
Biber, D. ii, 10, 26, 70, 113 n1, 132, 197–8, 204
Bolinger, D. 4, 33, 80, 112, 113
Bosworth, J. 154
Brinton, L. xii, xiii, 2, 5–6, 35, 44, 64, 68, 70–1, 130
Burnard, L. 174, 206 n2, 207 n8
Busse, U. 241–2, 249 n18
Buyssens, E. 102–3
Bybee, J. L. ix, 18–19, 26, 35–6, 57–8, 70–4, 122, 187, 189, 191, 230, 245

C
Campbell, L. ix, x, 27 n2, 211, 229, 247 n2
Cassidy, F. G. 249 n14
Chambers, J. K. 14
Christophersen, P. 108
Clarke, S. 223 n8
Claudi, U. 4, 154
Coates, J. 37–9, 43–5, 49, 59
Compes, I. 131–2
Croft, W. 193
Crowell, T. L. 35
Curme, G. O. 99, 101, 103, 105, 108, 110, 115 n22

D
Danchev, A. 128, 140 n4
Dasher, R. B. 70–72, 74 n4, 75 n5, 242
de Beaugrande, R. 204, 208
de Mönnink, I. 208 n24
Denison, D. 35, 39–40, 53 n4, 54 n10, 86–7, 93–6, 98–104, 108, 113–5, 138, 249 n16
Detges, U. 188
Di Micello, P. 175
Diewald, G. ix, 212
Donner, M. 19
Dyvik, H. 58

E
Edmondson, W. 60, 62
Ehram, M. E. 53 n3
Ellis, A. J. 218
Erades, P. A. 110

F
Facchinetti, R. 59–61
Finegan, E. 113 n1
Fischer, O. ix, 250 n23

G
Giacalone Ramat, A. ix
Gisborne, N. 7–9
Givón, T. 4, 87, 114 n14, 122
Görlach, M. 138
Greenberg, J. 189, 207 n12

H
Haiman, J. 186–7
Hall, J. H. 249 n14
Halliday, M. A. K. 3–5, 9, 99
Haspelmath, M. 181
Heikkonen, K. x
Heine, B. ix, xiv, 4, 27–8, 30, 54, 73, 78, 120, 135, 149, 150, 152, 154, 169, 172, 188, 207, 209, 210, 225–6, 253
Hogg 28–30, 231, 254
Hopper ix, 4, 5, 18–19, 34, 36, 71, 81–2, 86, 107, 113 n2, 122, 125–6, 128, 131, 152, 168 n3, 169, 172, 177–8, 180, 188, 213, 216, 225, 247 n1
Huddleston, R. 9, 132, 168 n1
Hundt, M. xi, 126, 140
Hünnemeyer, F. 4, 154

J
Janda, R. ix, x, 15, 27 n2, 228–230, 245, 247 n4
Jeffers, R. J. 228–230, 244, 247 n3
Jespersen, O. 9, 34, 90, 92, 94–5, 101, 110, 114 n9, 115 n24, 120, 128, 137, 249 n19
Johansson, S. 10, 12
Jones, C. 12
Jordan, R. 244, 250 n27
Joseph, B. 229, 247

K
Killie, K. 7–8, 17
Kjellmer, G. 74
Klinge, A. 59
Koivisto-Alanko, P. 168 n2
König, E. 141 n7, 172
Koopman, W. 231, 249–50 n23
Kortmann, B. 140–1 n7, 168 n2
Kroch, A. 13, 227–8, 237–8, 242, 246
Krug, M. xi, 33, 35–6, 40, 42, 52, 122, 126, 186–7, 189
Kryk-Kastovsky, B. 249 n20
Kuteva, T. 73, 152
Kutscher, S. 131
Kytö, M. x, 29, 123, 128, 140 n4, 168 n10, 230

L
Labov, W. 16, 47, 48, 50, 53, 219
Laing, M. 27 n5
Langacker, R. W. 180, 186
Lass, R. 214–6, 221, 242, 246
Lee, D. Y. W. 202, 207 n21
Leech, G. 59
Lehmann, C. 4, 131–2, 172, 199, 202, 206 n6, 209, 212, 227, 229–30, 244, 248 n4
Lessau, D. A. 228
Levinson, S. 62
Lightfoot, D. 126
Lindvall, L. 74
Locher, M. 206
Lowth, R. 20, 26
Lutz, A. 242

M
Macaulay, R. K. S. 26
Mair, C. xi, 172, 223
Marchand, H. 17, 29
Markus, M. 12
Matsumoto, Y. 229
Maynor, N. 219
McEnery, T. 10, 196
McWhinney, B. 27 n6
Menn, L. 27 n6
Meurman-Solin, A. 27 n5
Millward, C. 249 n16
Minkova, D. 247
Mishoe, M. 219, 221–2, 223 n7
Mitchell, B. 53 n4, 160
Mossé, F. 90, 95, 231, 232
Mukherjee, J. 206
Mulac, A. 122

Myhill, J. 228

N
Nehls, D. 98
Nespor, M. 243–4, 250 n25
Nevalainen, T. v, xi, 140, 214, 215
Nevis, J. A. 229
Newmeyer, F. xii, 138, 211, 227–9, 247
Noël, D. 58
Norde, M. 230, 245
Nurmi, A. 114 n15
Nuyts, J. 70

O
Oizumi, A. 231
Olofsson, A. 122
Österman, A. 160

P
Pagliuca, W. ix, 71, 230
Palander-Collin, M. 5
Palmer, F. R. 34, 38–9, 43, 53 n5, 59–62, 74
Panther, K. U. 60–1, 63
Paradis, C. 20
Paul, H. 185–6
Pawley, A. 186
Perkins, R. 230, 245
Pintzuk, S. 228
Plungian, V. A. 58–9, 65, 71
Poplack, S. 37
Pounder, A. 26, 27 n3
Pratt, L. 103, 108
Pullum, G. 9, 132, 168 n1
Pustejovsky, J. 64

Q
Quirk, R. 4, 6–9, 17, 23–5, 70, 99, 132, 168 n1, 173, 182, 199

R
Rand, D. 49
Raumolin-Brunberg, H. 2, 5, 11, 14–16, 214–5
Renouf, A. 17
Rickford, J. R. 139
Rissanen, M. x, xii, 1, 3, 4, 14, 20, 23–4, 27 n8, 122, 207 n19, 222, 237, 249 n16, 250 n29
Rosenbach, A. ix
Ross, J. R. 4, 214, 217, 220
Rubino, C. 229, 247 n2
Rudorf, C. 131

S
Sadock, J. 243
Sand, A. 114 n17
Sankoff, D. 49
Sapir, E. 3, 18
Scheffer, J. 93, 94, 101, 108
Schwenter, S. A. 172, 175–6
Siemund, R. 114 n17
Skandera, P. 114 n17
Smith, N. 100, 114 n18, 249 n13
Smitterberg, E. 109, 115 n23
Sobkowiak, W. 193
Stein, D. ix, 246
Stern, G. 71
Strang, B. 232, 248 n6
Sundby, B. 20
Svartvik, J. 86
Swan, T. 17, 24
Sweetser, E. E. 188
Syder, F. H. 186

T
Tagliamonte, S. xi, 140
Tarpent, M.-L. 247
Thornburg, L. 60–1, 63
Toller, T. N. 154
Tottie, G.. 140, 206
Trask, R. L. 228
Traugott, E. C. vii, ix, 4–5, 16, 18, 27 n2, 35–6, 41, 58, 70–2, 74, 75 n5, 113 n2, 125–6, 128, 131, 135, 152, 168 n3, 172, 175–6, 178, 180, 188, 213, 242, 247 n1
Trudgill, P. 42, 54 n11

V
Van der Auwera, J. 58–9, 65, 71, 212
van der Wurff 250 n23
van Kemenade, A. 237, 247, 249–50 n23
van Riemsdijk, H 249, 250 n26
Venezky, R. L. 231
Vennemann, T. 247
Viberg, Å. 65
Visser, F. T. 34–5, 40, 86–7, 89–92, 94, 95, 100–4, 108, 113 n4, 114 n13, 115 n24, 128, 137, 249 n16
von Wright, E. H. 59

W
Wakelin, M. 213
Waltereit, R. 188
Warner, A. 34–5, 40, 53 n3, 97–8, 100–3
Williamson, K. 27 n5
Willinsky, J. 140 n1
Wilson, A. 10, 196
Wischer, I. ix, 212
Wray 194
Wright, J. 248
Wright, L. xii

Y
Yang, L. 122

Z
Ziegeler, D. 102, 103
Zwicky, A. M. 228—— 30, 243, 244, 247 n3, 249 n22, 250 n26

Subject index

A

AAVE see African American Vernacular English
ability xi, 57–60, 62–5, 67, 71–4
absolute frequency 129, 171, 190, 192
abstractness 23, 151–5, 157–67, 169, 172, 177, 180–1, 185, 191, 195, 198
adverb xii, 2, 6–8, 10, 12, 17–20, 22–7, 151, 154, 156, 158, 160–3, 165–6, 169, 229
adverbial xii, 16, 18, 20–25, 90, 143, 153, 159–60, 165, 167–8
adverbialization 1–4, 16–17, 19–20, 23, 25, 27
African American Vernacular English xii, 211–13, 219–22
age of speaker 33, 37, 41–2, 47–52, 173
American English 34, 39, 52, 80, 98, 110–11, 114, 124, 133, 140, 166, 174, 194, 196, 202–3, 213, 219–22, 238–9, 249, 253
analogy xii, 171, 180, 183, 194–5, 200, 205
animacy 43–4, 47, 51, 87–8, 91, 100, 114, 157–8
ARCHER corpus 14, 21–2, 27, 79–80, 87–91, 93–100, 104–6, 108–9, 110–11, 113–119, 124–5, 152, 166, 217
aspect 57, 69–70, 73, 79–81, 83, 86, 94, 98–102, 106–7, 112–3, 158, 219–20, 222
attrition 178–179, 188, 245
auxiliary 9, 33, 39, 40, 42, 54, 62, 71, 73, 80, 96, 108, 129–30, 212, 237–8, 241, 243

B

bleaching 73, 141, 178, 188
BNC see British National Corpus
British National Corpus xiv, 36, 54, 84, 91–2, 98–9, 106, 109, 152, 167, 172–4, 176–8, 180–4, 191–4, 198, 200–207, 231, 240
Brown Corpus 110, 152

C

Century of Prose Corpus 152
cline 2, 4–6, 8–9, 11, 16–18, 27, 43, 245
clitic 5, 18, 227–250 passim
cliticization xiii, 187, 227–8, 237, 242, 245–6
communal change 16
complex preposition xii, 11, 172–206 passim
conceptual frequency xii, 185, 190–1, 198
concreteness 152, 154–5, 157–61, 163, 165–7, 169, 172, 175, 177–81
conjunct 151
conjunction 10–11, 127, 133, 137, 172, 207
connective 152, 168, 229
constant-rate-hypothesis 13
conventionalization 60, 72, 172, 179
copula 8
Corpus of Early English Correspondence 4, 14, 27, 30, 120, 214, 226
Corpus of Early English Correspondence Sampler 152, 169
Corpus of Middle English Prose and Verse 152, 163

D
decliticization xii, xiii, 227–50 passim
degrammaticalization xi, xiii, 6, 12, 212–23 passim, 227–28, 245, 247
delicacy 9, 204
demorphologization 229, 247
deontic xi, 33–53 passim, 57, 58, 65, 71, 73, 137
derivation 16–19
diachrony xi, xiii, 3, 4, 13, 20, 25, 41, 71, 80, 82, 84, 109, 123, 126–7, 133, 135, 137, 152, 173, 175–177, 179, 181, 191, 198–9, 207, 211, 242
Dictionary of Old English Corpus 152, 154–5
discourse frequency xii — xiii, 95, 105, 123, 125–27, 130, 138, 172
discourse marker 238–9, 242
drift 247

E
Early Modern English 5, 7, 14, 20–1, 23, 125, 136, 149, 152, 165, 169, 227–28, 237–49 passim
emphasizer 23–24
enclitic 229–49 passim
encliticization see enclitic
English-Swedish Parallel Corpus xi, 58, 75–7
epistemic 34–5, 38–9, 51, 57–8, 71, 73, 77, 122, 137
ergative 103, 106–7, 115
Estonian 229
exaptation 221

F
FLOB see Freiburg-Lancaster Oslo/Bergen Corpus
Freiburg-Lancaster Oslo/Bergen Corpus 22, 99–101, 109–10, 114, 118, 128, 140 152–3, 166–7
Freiburg-Brown Corpus 22, 99, 109–10, 114, 118, 128, 140, 152–3, 166
Frown see Freiburg-Brown Corpus

G
generalization xii, 15–16, 26, 125, 155–6, 177, 188
generational change 15
generic 33, 45–51, 50–4, 60, 105–6
gradience 2, 4, 9–11
grammatical indeterminacy 7
grammaticization 81
Greek 230
group genitive 230, 247–8
Gutenberg, Project 174–8, 191–2, 198, 206, 209

H
Helsinki Corpus of English Texts 14, 152, 170, 230
his genitive 230, 247
Hungarian 78, 229

I
Ilokano 229
inanimacy 44, 52, 59, 88, 91, 100, 154, 157–8
inflection 5, 16–19, 23, 25, 27
intensifier 23–24
intersective gradience 9

J
Japanese 229

L
Lampeter Corpus of Early Modern English Tracts 21
Lancaster-Oslo/Bergen Corpus 10, 22, 98–100, 109–10, 114, 118, 120, 128, 140, 152
Lapp 229
Latin 140, 164, 230, 251
layering xi, 19, 27, 35, 40, 42–3 51, 53, 79–84, 86, 92, 101–5, 107, 112–3, 177, 201, 206, 213, 216–7
levelling 12, 50, 52
lexicalization xi, 1, 3, 6, 16, 211–2
lexico-grammar xi, 1
LOB see Lancaster-Oslo/Bergen Corpus

low-frequency xii, 171–205 passim

M
manner adverbial 24
markedness xi, 50, 79, 82, 84, 91, 94, 105–7, 112–4, 221–2
maximizer 23–4
mediopassive 80, 87, 103, 105–107, 112, 115
Middle English xii, 10–12, 18, 19, 23, 27, 34–5, 44, 134, 151–168 passim, 212, 214, 227, 231–3, 237, 240, 248
Middle English Compendium 152, 163, 253
modality xi, 24, 33–77 passim
moderator 23–4
multivariate analysis 33, 47, 49–50

N
necessity xi, 33, 35, 40, 42–5, 50–3
normalcy 1, 2, 9–12, 204
normalization 12, 29

O
obligation xi, 33, 35, 40, 42–5, 50–3, 73
Oxford English Dictionary 23, 26, 35, 123–5, 155, 164, 175, 209, 231, 253

P
palatalization 244, 246
paradigmatic 5, 18, 113, 199
Parsed Penn-Helsinki Corpus of Middle English 12, 27
participant-external actuality 65
participant-internal actuality 65, 73–4
passive xi, 68, 75, 76, 79–118 passim, 122, 137, 140
periphrastic *do* 97, 117, 228
permission xi, 34–5, 53, 67, 71–2, 74, 161
phonogenesis 247
portmanteau tag 10, 12
possibility xi, 33, 34, 57–9, 61, 67, 71–4
PPCME see Parsed Penn-Helsinki Corpus of Middle English
pragmatic implicature 60
pragmatic inferencing 172
preposition xii, 10–11, 122, 151–206 passim
proclitic 228, 236
pronominalization 4
Proto-Germanic 231
Proto-Indo-European 230

Q
quasi-modal 34, 39

R
replacement xiii, 14–16, 103, 207, 212, 214, 227, 242–3, 245–7
representativeness 173, 197, 198, 204–5

S
saliency 185, 189, 191, 193, 195, 205
sandhi 244–5, 247
Sanskrit 230
semantic change 23, 72, 125, 177, 179
serial relationship 4, 30
sex of speaker 33, 37, 48, 173
social embedding 2, 4–5, 14–15, 30
sociolinguistics xi, 1–2, 4, 13, 14
Spanish 153, 230
specialization v, 33, 36, 43, 50, 52, 80, 82–83, 91, 103–107, 167
squish 4
subject enclisis xiii, 228, 238, 243
subjectification xii, 151–2, 162, 170, 178
subjunctive 213, 217–8, 220, 222–4
subordinator xii, 14, 136, 151, 165, 169, 217, 220, 223
subsective gradience 9
success 58, 62–3, 67, 71–4
Swedish xi, 57–77 passim, 230
synchrony 3
syntagmatic 2, 5–8, 72, 187, 199

T
total accountability 1, 2, 9–10, 12, 15
translation 57–77 passim, 140

U
unidirectionality xii, 4, 227, 247
univerbation 10, 227

V
variable rule 48–9, 50–1, 55

W
word-formation 3, 6, 8

Z
zero-formation 19, 23

In the series *Studies in Corpus Linguistics (SCL)* the following titles have been published thus far or are scheduled for publication:

1 **PEARSON, Jennifer:** Terms in Context. 1998. xii, 246 pp.

2 **PARTINGTON, Alan:** Patterns and Meanings. Using corpora for English language research and teaching. 1998. x, 158 pp.

3 **BOTLEY, Simon and Anthony Mark McENERY (eds.):** Corpus-based and Computational Approaches to Discourse Anaphora. 2000. vi, 258 pp.

4 **HUNSTON, Susan and Gill FRANCIS:** Pattern Grammar. A corpus-driven approach to the lexical grammar of English. 2000. xiv, 288 pp.

5 **GHADESSY, Mohsen, Alex HENRY and Robert L. ROSEBERRY (eds.):** Small Corpus Studies and ELT. Theory and practice. 2001. xxiv, 420 pp.

6 **TOGNINI-BONELLI, Elena:** Corpus Linguistics at Work. 2001. xii, 224 pp.

7 **ALTENBERG, Bengt and Sylviane GRANGER (eds.):** Lexis in Contrast. Corpus-based approaches. 2002. x, 339 pp.

8 **STENSTRÖM, Anna-Brita, Gisle ANDERSEN and Ingrid Kristine HASUND:** Trends in Teenage Talk. Corpus compilation, analysis and findings. 2002. xii, 229 pp.

9 **REPPEN, Randi, Susan M. FITZMAURICE and Douglas BIBER (eds.):** Using Corpora to Explore Linguistic Variation. 2002. xii, 275 pp.

10 **AIJMER, Karin:** English Discourse Particles. Evidence from a corpus. 2002. xvi, 299 pp.

11 **BARNBROOK, Geoff:** Defining Language. A local grammar of definition sentences. 2002. xvi, 281 pp.

12 **SINCLAIR, John McH. (ed.):** How to Use Corpora in Language Teaching. 2004. viii, 308 pp.

13 **LINDQUIST, Hans and Christian MAIR (eds.):** Corpus Approaches to Grammaticalization in English. 2004. xiv, 256 pp. + index.

14 **NESSELHAUF, Nadja:** Collocations in a Learner Corpus. *Expected Fall 2004*