English Language and Linguistics 6.1: 105–31. © Cambridge University Press 2002
DOI: 10.1017/S1360674302001065 Printed in the United Kingdom

Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora

CHRISTIAN MAIR
University of Freiburg
(Received 12 July 2001; revised 2 November 2001)

The article looks at three instances of grammatical variation in present-day standard English: the use of bare and *to*-infinitives with the verb *help*, the presence or absence of the preposition/complementizer *from* before *-ing*-complements depending on *prevent*, and the choice between *-ing-* and infinitival complements after the verbs *begin* and *start*. In all three instances, current British and American usage will be shown to differ, and these differences need to be interpreted against diachronic changes affecting Late Modern English grammar as a whole. The description of twentieth-century developments is mainly based on data obtained from matching corpora of British and American standard English. Since in all three cases studied developments did not originate in the twentieth century, additional data from the quotation base of the *OED* were used to outline the long-term evolution of the relevant portions of the grammar since ca. 1600. In general/methodological terms, the article aims to show that an utterance-based model of language change, in combination with the exceptionally well-developed corpus-linguistic working environment available to the student of standard English, can lead to new discoveries even in a well-studied area such as the grammar of standard English.

1 Introduction: the Freiburg Project

When, with the advent of computers, corpus studies got a new lease of life in linguistics in the 1960s and 1970s, two matching one-million-word samples of written British and American English, the LOB and Brown corpora, provided much of the focus for this type of computer-assisted descriptive work. By the early 1990s an impressive body of studies on British–American contrasts had emerged based on these two collections of texts.[1] The obvious success of LOB and Brown inspired similar ventures, in which matching corpora were added for second-language Indian English (the Kolhapur Corpus), Australian English (the Macquarie Corpus) and New Zealand English (the Wellington Corpus). All these resembled the original Brown and LOB corpora in their size and composition, but unfortunately no longer in their sampling dates, which had been 1961 for texts included in LOB and Brown but ranged from the late 1970s to the late 1980s for the remaining corpora. It is thus obvious that an unwelcome diachronic distorting factor was introduced into what was envisaged as a synchronic record of regional variation in standard Englishes.

[1] Cf., e.g. the bibliography in Altenberg (1991). The corpora are named after the institutions responsible for their compilation: the universities of Lancaster (UK) and Oslo/Bergen for LOB, and Brown University for Brown. For a brief history of corpus linguistics, including basic information on composition and availability of the major databases mentioned here and further on in this study, see McEnery & Wilson (2001) and Taylor, Leech & Fligelstone (1991).

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Addressing this problem, the Freiburg-based corpus-linguistic research group decided to replicate the original pair and eventually compiled the updates of LOB and Brown commonly referred to as F-LOB (Freiburg update of LOB) and Frown (Freiburg update of Brown). The sampling year was 1991 for F-LOB and 1992 for Frown, which broadly corresponds to the interval of one generation that is usually considered the minimum period required to clearly identify and document linguistic change in real time.[2] Like the corpora mentioned above, F-LOB and Frown are available to interested members of the linguistic community through ICAME, the International Computer Archive of Modern/Medieval English (http://www.hit.uib.no/icame).

In addition to the immediate aim – providing up-to-date British and American material for comparison with the more recent Australian and New Zealand corpora and thus extending the lifespan of work on regional variation of the type envisaged by the compilers of LOB and Brown, – a quartet of matching corpora covering British and American English thirty years apart provides an opportunity to demonstrate the appropriateness of the time-honoured sociolinguistic stipulation that the study of diachronic change and synchronic (regional, stylistic, social) variation should be integrated for a better understanding of either phenomenon.

The following are possible goals of research based on these corpora; they are listed in increasing order of ambitiousness. The goals of the project are threefold:

1. To empirically verify/falsify those hypotheses on linguistic change in present-day English which are proposed in the linguistic literature. Such testing seems particularly necessary in view of the fact that most such hypotheses are based on anecdotal observations by individual commentators.
2. To uncover instances of change and/or variation not previously noticed in the literature through a systematic and exhaustive comparison of frequencies in the corpora.
3. To use recent developments in the British and American written standards in order to investigate the precise mode of interaction between synchronic variation and diachronic change.

The corpora being relatively small (one million words each) and documenting edited written English, work on them has taken a natural direction towards the analysis of the core grammatical structures of English and certain medium- to high-frequency collocational phenomena at the interface of lexicon and grammar. This restriction in

[2] Writing on phonetic change, Labov says that confirmation of a suspected linguistic change is obtained 'if it is demonstrated in the near future that the trend detected has moved further in the same direction. "Recent past" and "near future" must mean a span of time large enough to allow for significant changes but small enough to rule out the possibility of reversals and retrograde movements: we might say from a minimum of a half generation to a maximum of two' (1981: 177). For lexical change, the minimum span of observation may even be shorter. Syntactic change is generally agreed to proceed more slowly than either lexical or phonetic change. But even if the whole life-cycle of an innovation – from the origin of a new structure to its full integration in the core grammar – takes centuries, thirty years may nevertheless mark a significant step in the spread of an innovation.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

scope is not at all unfortunate, because it yields results on precisely those aspects of change in present-day standard English for which reliable documentation is most difficult to obtain.

2 Observing grammatical change in progress

Even surface-oriented or utterance-based models disagree as to whether the direct observation of ongoing linguistic change is possible at all. The *locus classicus* of the skepticist position is probably the following remark in Bloomfield's *Language*:

> The process of linguistic change has never been directly observed; we shall see that such observation, with our present facilities, is inconceivable. (Bloomfield, 1933: 347)

It could be argued that while Bloomfield here speaks of linguistic change in general, the context of the passage makes clear that his actual topic is phonetic change. However, a subsequent comment on the rise of new analogical plurals (1933: 408) suggests that Bloomfield was similarly pessimistic about the possibilities of directly observing grammatical or lexical innovation. This skepticism contrasts sharply with the brash optimism expressed in the introduction to Bauer's *Watching English change*:

> This book will show that English is changing today and that you can watch the changes happening around you. (Bauer, 1994: 1)

On the face of it, the two positions are incompatible. However, interpretation of the passages quoted in their full context leads to some qualifications, and suggests possible compromises between the apparently contradictory views. Elsewhere in his work Bloomfield, for example, points out:

> *Fluctuation in the frequency of speech forms* is a factor in all non-phonetic changes. This fluctuation can be observed, to some extent, both at first hand and in our written records. (1933: 393; emphasis in the original)

Moreover, between the days of classical American structuralism and the digital age progress has been impressive in corpus linguistics, both in terms of quantity of material available for analysis and the sophistication of recording, storage, and retrieval techniques. 'Our present facilities' (to take up the formulation used by Bloomfield, 1933: 347, in his hedge) may thus have improved to an extent that might make Bloomfield rethink his earlier position. The main body of Bauer's book, on the other hand, is an extended demonstration not only of the potential of his corpus-based approach to language change, but also of its many limitations, and certainly does not fully endorse the breezy optimism apparent in the introductory remark quoted.

Subsequent developments in linguistics, however, have proved Bloomfield wrong on one detail. Owing to the work of Labov (e.g. 1994) and other variationists, we now know a fair deal about ongoing phonetic change (i.e. the area which Bloomfield was categorically pessimistic about). By contrast, hardly anything is known even

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

today about ongoing changes in morphology and syntax (i.e. those areas in which Bloomfield saw most opportunities). Part of the explanation may be that the study of syntactic change until recently was dominated by 'abstract' or 'formal' rather than 'contextual' or 'utterance-based' models, and abstract models generally privilege the long time range and the underlying system at the expense of short-term fluctuation and the statistical analysis of large masses of authentic utterances (cf. e.g. Lightfoot, 1991). A focus on gradual, incomplete and ongoing changes is more compatible with utterance-based models such as grammaticalization theory; but even here one notes that the emphasis has tended to be on remote and completed changes (e.g. the emergence of the *going to* future) rather than ongoing and incomplete ones.

The same order of priorities apparent in technical linguistic studies of language change is evident also in more popular works on the recent history of English. Standard accounts of change in present-day English such as Barber (1964) or Potter (1975) devote far more space to phonological and lexical phenomena than to grammatical ones. What is worse, the quality of the information available on grammatical matters tends to be inferior, too. The reason for this is simple enough. Outside sociolinguistic and variationist studies on change in progress (which are usually carried out on nonstandard varieties rather than the standard), the chief methodology employed is anecdotal observation, which, assuming an informed and competent observer, may work rather well with lexical innovation, possibly even with phonetic change, but is clearly deficient when it comes to grammatical change, where the available literature is full of contradictions. Some sources give the impression that grammatical change in standard English has virtually come to a halt (cf. e.g. Greenbaum, 1986: 6f., who lists just one very minor example of changing article usage), while others have long lists of structures that are apparently undergoing change at the moment (e.g. Barber, 1964: 130–44). Barbara Strang points out one possible reason for this uncertainty about the scope of ongoing grammatical change:

> One possible explanation can hardly be proved false, but should be entertained only as a last resort: namely, that although there has been considerable grammatical change in the past, English grammar in our own lifetime is somehow uniquely stable and free from change.
>
> The most promising direction of search for an explanation would seem to lie in the assumption that there is grammatical change in progress at the moment, as in the past, but that we are considerably less perceptive of it than of other kinds of linguistic change. (1970: 59f.)

Why should we be less perceptive of it, though? The answer is that, at the surface level of utterances, texts, and discourse, grammatical change manifests itself indirectly – in shifting statistical distributions of constructional variants or potentially re-analysable constructions. The situation is summarized succinctly by Greenbaum:

> Over a period of fifty or so years, grammatical change manifests itself largely in the increased frequency of some variants over others, in stylistic restrictions on some variants, and in differences in the grammatical treatment of individual words. These

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

> changes spread gradually across the whole speech community, sometimes taking several generations before they become conspicuous. (Greenbaum, 1986: 6)

Clearly then, studying grammatical change in progress does not mean a futile hunt for the earliest attestation of a new construction. Rather, it means documenting its gradual spread, and to this end there is no better resource to start on than a set of matching representative corpora that can be exhaustively analysed.

Given the fact that the corpora used in the present study are limited in size and restricted to written English, the results will often be provisional. In this situation, it is a definite advantage that the student of standard English has access to the best corpus-linguistic working environment of any language. 'Close-range' results obtained from LOB, Brown, F-LOB and Frown can be tested on larger and more varied synchronic corpora of English, and also against an ever fuller range of diachronic resources.

Ultimately, though, corpus-based empiricism must not lose touch with the theoretical linguistic tradition in the study of linguistic change. If it did, it would degenerate into data-driven positivism with counting as its only methodology. What should be aimed at is the type of synthesis envisaged by Rickford and his co-authors in a paper on a change in progress in modern American English:

> This exploration on the boundaries of sociolinguistic variation, corpus linguistics, historical linguistics, and syntax demonstrates the value of bridging the gaps between subfields . . . the field could benefit from more such collaboration.
> (Rickford, Mendoza-Denton, Wasow & Espinosa, 1995: 128)

3 Brown/LOB/Frown/F-LOB: a typology of contrasts

In the following, I shall survey eight possible abstract numerical constellations that come up in the comparison of the four corpora, commenting and illustrating them briefly where necessary.

Constellations 1 and 8, while among the most frequent ones encountered, are of course the least interesting, representing two opposite types of regional and diachronic stability with regard to the grammatical feature under review.

Constellation 2 is largely self-explanatory. In 2, the '+/−' signs indicate that a variable occurs at a different (greater or lesser) but constant rate of frequency in the American samples. Of course, if for entirely justifiable reasons American English were taken as the standard of comparison, they would have to appear in the British column. For equally arbitrary reasons the earlier state was taken as the point of reference in constellation 3, and the British English state in 1961 as the point of departure in constellation 8.

Typical examples of stable regional contrasts are provided by the major orthographical contrasts between the two varieties (e.g. *-or* vs. *-our*, or *-er* vs. *-re*) and by some morphogrammatical contrasts of long standing, for example the variation between the past-participle forms *got* and *gotten*, or – to use a less widely known case – *towards* (a statistical Briticism) vs. *toward* (the US preference).

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

1. 'nothing happening'

	BrE	AmE
1961	=	=
1991/92	=	=

2. 'stable regional contrast'

	BrE	AmE
1961	=	+/–
1991/92	=	+/–

3. 'parallel diachronic development'

	BrE	AmE
1961	=	=
1991/92	+/–	+/–

4. 'convergence: Americanization'

	BrE	AmE
1961	+/–	=
1991/92	=	=

5. 'convergence: "Britishization" '

	BrE	AmE
1961	=	+/–
1991/92	=	=

6. 'incipient divergence: British English innovating'

	BrE	AmE
1961	=	=
1991/92	+/–	=

7. 'incipient divergence: American English innovating'

	BrE	AmE
1961	=	=
1991/92	=	+/–

8. 'random fluctuation'

	BrE	AmE
1961	=	+/–
1991/92	+/–	+/–

Figure 1. Eight statistical constellations encountered in the comparison of results from the four corpora

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Constellation 5 seems to be as marginal as the awkward term coined to refer to it – 'Britishization'. With a few arguable exceptions, such as the modal idiom *have got to*, whose spread may indeed be being spearheaded by British English, there is hardly any evidence that the grammar of international standard English is converging on current British norms. There is thus no need to heed the mock-alarmist call sounded in Algeo's 1990 paper titled 'The Briticisms are coming!' – at least as far as grammar is concerned.

This leaves us with constellations 6, 7, and 4. They will be illustrated in detail with three different types of nonfinite complement structures, respectively: namely (1) gerunds dependent on *prevent* (and a number of semantically related verbs), where there is variable use of a preposition/complementizer *from*, (2) complements of the aspectual verbs *begin* and *start*, where there is a choice between infinitives and gerunds, and (3) infinitival constructions with *help*, where there is a choice between *to*-infinitives and bare infinitives.

These phenomena have been singled out for analysis because they all represent instances of grammatical change in progress in Late Modern English, and have resulted in – sometimes temporary – regional contrasts between British and American English. Although all three phenomena have been studied and commented on extensively, fully convincing accounts are still lacking – mainly because analysts have failed to provide an integrated description doing justice to the interaction of synchronic variation and diachronic development. To give an example: to the investigator limiting himself/herself to a strictly synchronic perspective, the choice between bare and *to*-infinitives with *help* may well appear as a matter of British or American regional preference. It is only after one also takes into account diachronic developments that one will find that the more important phenomenon is in fact a general increase in the frequency of the bare infinitive in *all* varieties. Any observed regional contrast is thus temporary and due to varieties developing towards a common goal at slightly differential speeds. Needless to say, any generalization on semantic or stylistic values associated with the presence or absence of *to within* individual varieties will rest on even more precarious foundations, as in a situation of ongoing change an analyst with a synchronic orientation is in the position of someone shooting at a moving target.

It is a welcome side-effect of the choice of examples that they will show that there is more to British–American grammatical differences than well-established but hopelessly overworked standard examples such as the *got*/*gotten*-distinction or the mandative subjunctive. They all bear out a suspicion voiced by Algeo:

> Perhaps the most fertile area of divergence between British and American is the complementation of verbs. (Algeo, 1988: 22)

The three variable constructions studied here represent only a small fraction of a potentially far greater range of relevant contrasts. It is thus hoped that the present study will inspire more corpus-based work in the synchronic variability and recent history of standard Englishes around the world.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table 1. *Complementation of* prevent *in four corpora**

ratio of *prevent NP from V-ing* vs. *prevent NP V-ing*	BrE	AmE
1961	34:7	47:0
1991/92	24:24	36:1

(BrE diachr. $p<0.01$; all other contrasts not significant)

* One of the seven instances of *prevent NP V-ing* in LOB has *her* as the notional subject of the gerund and could thus have been excluded as representing the 'archaic' type (*prevent my leaving*) disregarded here. The sole American attestation of the 'British' pattern (in Frown) is from a work of military history dealing with, significantly, the Battle of Britain.

4 Incipient innovation in British English: *prevent* + NP + V-*ing*

An analysis of the four corpora shows at least one clear instance of constellation 6 (divergence, with British English innovating). The construction in question is best illustrated in the verb *prevent*, which in the relevant sense is constructed either with an archaic gerundial phrase (*this prevented my leaving early*), with the preposition/complementizer *from* (*this prevented me from leaving early*), or with a fused participle (i.e. a noun phrase in the object/common case followed by V-*ing*: *this prevented me leaving early*). The first, archaic type of gerund, with the notional subject in the genitive, is attested very rarely in present-day English and will be disregarded in the analysis. The figures for the remaining two constructions are given in table 1.

The results from our four corpora tally well with the figures from many other twentieth-century corpora and electronic databases, for example 1990s newspapers on CD-ROM. The *Guardian* (1992) shows the roughly even split between the two alternatives found in F-LOB, whereas the *Miami Herald* for the same year did not produce a single instance of the 'British' pattern for the first one-hundred relevant cases. The two major sources for spoken British English reveal patterns of variation practically identical to those obtained from the written ones. In the more conservative London–Lund Corpus, with its bias towards middle- and upper-middle-class speakers, there are 6 instances of *prevent from* as against 5 of *prevent* NP V-*ing* (in half a million words); the spoken-demographic component of the British National Corpus, with its more recent material and broader social coverage of informants, has 6 each (in a sample comprising ca. 4 million words). In the Corpus of Spoken Professional American English (Barlow, 1998, 2+ million words), by contrast, there are 35 instances of *prevent from*, and none of the other pattern.

That we are dealing with a grammatical phenomenon and not just with an idiosyncratic lexical property of one verb is shown by the behaviour of various synonyms of *prevent* such as *stop* or *block,* which are increasingly attested in *from*-less constructions in British English (although only *stop* occurs frequently enough for the trend to show up in our matching corpora, as shown in table 2).[3]

[3] The traditional construction with *from* is rare, but evenly spread throughout all four corpora. Brown additionally contains two 'archaic' constructions, with the noun phrase preceding V-*ing* in the genitive.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table 2. Stop NP V-ing *in four corpora*

	BrE	AmE
1961	4	0
1991/92	17	0

Occasional finds like the following, while insufficient to postulate a trend, further attest to the productivity of the *from*-less pattern in British English:

(1) Then, even though the EU president is supposed to be politically neutral, Prodi has declared that his new EU job and his new Democratic political party are 'two sides of the same coin'. Though the EU blocked him standing as a candidate in EU elections, he continues to participate in political campaigning.
(*Private Eye*, 30 April 1999, p. 16)

(2) Back at the top I helped Duncan carry the police barrier across the road so that it blocked traffic coming up. (BNC HWL 500)

Note that while the example (2) allows an alternative structural analysis – with the V-*ing* form functioning as a nonfinite postmodification of the preceding noun *traffic* – this option is foreclosed in the example (1), because the personal pronoun *him* is not a suitable head for a restrictive relative clause.

Further verbs which would merit attention in the present connection include *debar, deter, dissuade, discourage, free, hinder, prohibit.* They are all attested in V + NP *from* V-*ing* patterns. On the assumption that the core verbs *prevent* and *stop* are spearheading a grammatical change proceeding by lexical diffusion, they should tend towards the *from*-less pattern to varying degrees in British English. As they are too rare in the four corpora providing the starting point for the present investigation, they will not be treated here.

A somewhat complicated case is presented by *save* and *spare*, with which there is variable use of *from* particularly in British English, too. Since, in contrast to the *prevent*-type verbs, these two allow ditransitive complementation with nominal objects (cf. *save/spare someone the trouble* as against **prevent/stop someone the trouble*), V + NP + V-*ing* patterns need not necessarily be seen as reduced variants of V + NP *from* V-*ing*.

There have been several attempts to account for the variation observed in structural and/or semantic, and chiefly synchronic, terms. In a number of publications, Rohdenburg (e.g. 1995b, 1996) has invoked a 'complexity principle' to account for the distribution of the two types of complementation. The 'explicit' *from*-complement is said to be preferred with syntactically more complex environments (for example passives or long and complex object NPs) and to have spread diachronically (1995b: 87). These claims, however, are not entirely consistent with the data. Historically, there has been a spread of the *from*-complements, but only in American English, where they have ousted their 'bare' alternatives; in British

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

English the recent decades have been characterized by a spread of what according to Rohdenburg would be the dysfunctional (less explicit) alternative. Synchronically, the complexity principle may account for some of the variation observed in British English; in American English, where there has ceased to be variation, the principle cannot be applied any longer.

Dixon sees iconic factors at work in the variation between the two constructional patterns:

> *I prevented her going* would be likely to be used when I employed some direct means, e.g. I blocked her path. In contrast, *I prevented her from going* would be the appropriate thing to say if I used indirect means, e.g. I used my influence to make sure she didn't get her passport renewed. (Dixon, 1991: 236f.)

As with Rohdenburg's functional explanation, this account tallies with part of the attested data only. There are straightforward counterexamples, and in many other cases the semantic distinction involved is irrelevant or difficult to perceive in a given context.[4]

Considering that we are dealing with an obvious grammatical difference between the two varieties, the literature on British–American grammatical contrasts is not very enlightening on the subject, either. In many relevant lists, *prevent* is not mentioned as a variable at all (cf., e.g., Hansen et al., 1996: 128–30); in others (e.g. Trudgill & Hannah, 1982: 56) it is mentioned under a heading – prepositional usage – where it arguably does not belong. Algeo (1988: 24) notes the absence from American English of *from*-less complements after *stop*, but then goes on to say that the corresponding structure with *prevent* is accepted 'by many, though not all, American consultants'. This is difficult to reconcile with the absence of attestations from recent American English, but understandable in view of the fact (cf. below) that the construction was apparently used fairly widely in American English in the past. It is only Dixon (1995: 217) who, in another context, correctly notes that *from* can be omitted after *prevent*, *stop*, *save*, and *spare* in British and Australian English, but not in American English.

Part of the confusion is certainly due to the fact that this particular contrast between British and American English is relatively recent. The earliest variant is the construction with the nominal gerund (which is not considered here); the two variants with verbal gerunds, by contrast, are later and seem to have spread more or less simultaneously in the course of the eighteenth century (see, e.g., the examples in Jespersen, 1909–49, V: 149). The variant with *from* was probably patterned on an older use of *hinder from*, a construction normal since Middle English times (Visser, 1970–8: III: 2371), and the alternative without *from* arose in the context of the

[4] Cf., e.g., *the man who prevented him reaching Downing Street* (FLOB B12 121–2), where the reference is clearly not to physical restraint but indirect political machinations. The distinction is also impossible to apply in the numerous cases in which the NP following *prevent* is inanimate, abstract, or represented by the demonstrative pronouns *this* or *that*, as in *prevent our picture of the Tunisian campaign becoming disbalanced in favour of America* (FLOB G49 172–3).

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

systematic structural ambiguity between genitive/possessive and object/common case readings of certain pronouns and plurals at the time (see Fanego, 1998: 92, 101).[5]

While, as far as it is possible to tell, the early examples of *from*-less gerunds collected by Visser, Jespersen, and the OED are all British, it is not difficult to find eighteenth-century American instances, which shows that the same type of variation found in modern British English must at one stage also have been characteristic of American usage. Consider the following cases found in an electronic edition of Benjamin Franklin's *Diaries*, which is part of *The American Library: Wordcruncher Disc* vol I:[6]

(3) . . . every Law we have hitherto made to prevent this Deluge of Wickedness overwhelming us. (Franklin 1736–57, 359: 1)
(4) So that, while the Enemy is in the Heart of the Country, Cavils prevent any Thing being done for its Relief.
(Franklin 1757–75, 507: 1, repeated in slightly altered form 510: 2)
(5) There is no test to prevent Churchmen holding offices. (ibid., 676: 4)

Such (and similar) cases also explain why – against all present-day evidence – the *Columbia guide to English usage* (Wilson, 1993) claims *prevent* to be 'standard in all constructions' (343), and *Webster's third* lists the *from*-less variant before the one with *from*. Webster's *Dictionary of English usage*, by contrast, has got it almost right in claiming that the *from*-less form is 'fairly uncommon in American English, but quite common in recent British usage' (1989: 770).

What distinguishes the eighteenth-century variation from the present state of affairs is the rapid spread of the *from*-less variant in British English, which is without a parallel in the American variety. This is a development which, if carried to completion, will mean that from a common stock in which both structural options were present, British English standardized one, and American English the other. Alternatively, the dynamic apparent in twentieth-century British English may subside, and the British variety may continue to have two options where there is only one in American English. Whichever analysis ultimately turns out to be the correct one, though, one thing is certain: we would not have been able even to state the facts without the analysis of a quartet of matching corpora.

5 Gerundial complements after *begin* and *start*

The development to be dealt with in this section is in many ways a mirror image of the preceding one, illustrating constellation 7, incipient divergence in American

[5] Visser (1970–8, III: 2352) has early examples of the *from*-less variant going back to the year 1592, and the OED (s.v. *prevent*) has one instance from 1689. However, as Teresa Fanego points out (personal communication), none of these early examples are truly diagnostic because of the fluctuation in the use of the apostrophe with genitives at the time and the ambiguity of the pronoun *her* (possessive and objective case), which both make it impossible to decide whether we are dealing with nominal ('archaic') or verbal ('modern') gerunds.

[6] Orem, UT: Electronic Text Corp., 1990. References are to this edition.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

English. In contrast to the preceding example (*prevent*), the various structural and semantic factors favouring or requiring the use of infinitives and gerunds after *begin* and *start* are well described in reference grammars and the specialist linguistic literature (cf. e.g. Quirk, Greenbaum, Leech & Svartvik, 1985: 1191–3; Rericha, 1987; Schmid, 1993; Duffley, 1999). These sources agree that, in spite of a number of strict structural constraints and some fairly strong semantic tendencies favouring the infinitive,[7] considerable scope remains for stylistic variation.

The complementation of *begin* and *start* has also been a much studied issue in the corpus-linguistic community. In Biber, Conrad & Reppen's (1998) introduction to corpus linguistics 'the grammatical associations of nearly synonymous words: *begin* vs. *start*' serve as a pedagogical study example (95–100). The main insight of this pilot study is that 'relative to *start, begin* has a greater preference for the *to*-clause pattern' (99). This is correct but does not take into account additional differences between British and American usage which will be discussed below. The subject is taken up again in Biber et al. (1999: 373), where the authors observe that *start* is informal, occurring at a rate > 500/million words in the conversational subcorpus (the corresponding figure for *begin* is < 200/million). Concentrating on uses of *begin* and *start* with gerunds, Biber et al. provide further instructive statistics on p. 747 (figure 9.25). *Start* + V-*ing* outnumbers *begin* + V-*ing* in all four registers compared (conversation, fiction, news, academic prose), although the proportional and absolute frequencies vary drastically. Another corpus-based analysis is Mindt (2000: 322–5, 333–5), who corroborates Biber, Conrad & Reppen on the basic point but proceeds to undertake a more fine-grained analysis of usage in different registers. *Start* + V-*ing* is least frequent in fiction (34 per cent), more frequent in expository prose (48 per cent) and most frequent in conversation (71 per cent). The corresponding figures for *begin* are 11 per cent, 17 per cent and 4 per cent respectively.

All these accounts, whether system-based/descriptive or utterance-based/variationist, however, remain incomplete because they do not take into account that, as in the previous example, we are dealing with an instance of change in progress. With many classes of verbs, -*ing*-complements have gradually been spreading at the expense of infinitives since Early Modern English. Tables 3 and 4 give the proportions of *to*- vs. -*ing*-complements after *begin* and *start* in Brown, LOB, Frown, and F-LOB, and show the extent of the diachronic change in the latter part of the twentieth century.

As these tables make clear, there has been near stability with regard to the complementation of *start* in the period under review. Both infinitives and gerunds were, and continue to be, commonly used with this verb, and even the increase in the frequency of gerunds in the recent American material is only weakly significant.

It is the complementation of *begin* that we have to turn to for language-historical and regional 'action'. The corpora show there to have been a persistent regional

[7] To mention the two most important ones: -*ing*-complements are not attested after *beginning* and *starting* (* *I'm beginning working*), and they are strongly disfavoured when the verb is stative (**only then did I begin understanding what she meant*).

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table 3. to-*inf. vs. V*-ing *after* begin

	BrE	AmE
1961	260:23	230:53
1991/92	204:20	202:95

(BrE vs. AmE 1961 $p<0.001$; BrE vs. AmE 1991/92 $p<0.001$, BrE diachr. not significant, AmE diachr. $p<0.001$)

Table 4. to-*inf. vs. V*-ing *after* start

	BrE	AmE
1961	36:52	47:49
1991/92	49:59	59:110

(AmE diachr. $p<0.05$, all others not significant)

Table 5. to-*inf. and V*-ing *after* begin *in selected genres*

	Brown	Frown
A-C (press)	22:10	19:26
D-J (other nonfiction)	126:25	88:37
K-R (fiction)	82:18	95:32

contrast *between* British and American usage, on the one hand, and a highly significant diachronic development *within* American English, on the other. In British English, the gerund was infrequent in 1961, and has remained so. In American English, the gerund was more frequent than in British English even back in 1961 and has increased further since.

As is not surprising in the early stages of change in progress, this increase has not proceeded across the board, but at differential speeds in different textual genres. As table 5 shows, the dramatic increase of gerunds in the press sections of Frown is not paralleled in the remainder of the material.

The full complexity of the situation becomes apparent if one leaves the small and tidy corpora behind and ventures a few forays into dirtier data, as I did in consulting the array of corpora and electronic full-text databases shown in table 6. They provide additional evidence for written usage and some important facts on the situation in spoken English. Note that the large mass of newspaper material available makes it practical to also compare the frequencies of gerunds and infinitives after individual forms of *begin*, such as *began* or *begins*, to check whether the tense or morphological form of the matrix verb might have an influence on complementation. This factor cannot be investigated systematically in the smaller corpora.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table 6. *Complementation of* begin *in selected further corpora and electronic text-samples*

	to-inf.	gerund
first 100 relevant instances from *Los Angeles Times* 1992	55	45
first 100 relevant instances from *Boston Globe* 1992	57	43
first 100 relevant instances from *Miami Herald* 1992	24	76
first 100 relevant instances from *Guardian* 1996	86	14
first fifty instances of *began*/*LA Times* 92	21	29
first fifty instances of *began*/*Boston Globe* 92	19	31
first fifty instances of *began*/*Miami Herald* 92	10	40
first fifty instances of *began*/*Guardian* 96	38	12
first fifty instances of *begins*/*LA Times* 92	34	16
first fifty instances of *begins*/*Boston Globe* 92	38	12
first fifty instances of *begins*/*Miami Herald* 92	30	20
first fifty instances of *begins*/*Guardian* 96	48	2
Faulkner, *Light in August*	157	2
Cather, *O Pioneers!*	51	6
Thoreau, *Walden*	48	1
Whitman, *Complete Prose*	42	1
Emerson, *Essays*	67	–
London-Lund Corpus of Spoken British English	52	5
British National Corpus/Spoken-Demographic Sample	79	8
Corpus of Professional Spoken American English	288	40

The British data display the type of internal consistency one would expect in the absence of diachronic change. All figures from the *Guardian* show that the infinitive is the statistical norm, and the V-*ing* complement a relatively rare additional option. This is in line with the results from LOB and F-LOB and, what is more important, with the results from the two corpora of spoken British English. The choice of complement after *begin* is thus clearly not a style marker in British English separating written and spoken English, nor is it likely to be a sociolinguistic variable. If it was, some difference in distribution would be expected in spoken samples such as the BNC and LLC, which after all are extremely different with regard to the type of informant whose usage they document.

While we find stability in British usage, we see massive variability *within* American English. Frequencies range from ones that are rather close to what is found for British English in the Corpus of Spoken Professional American English[8] to very

[8] Pending the completion and public availability of the long-awaited Santa Barbara Corpus of Spoken American English, this corpus will continue to be the best resource for the study of spoken American English in spite of its bias towards the language of press conferences and public hearings. In the fourteen texts of the Santa Barbara Corpus currently available through the Linguistic Data Consortium, there

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

high frequencies of gerunds in narrative-past contexts in newspapers (cf. the figures for *began*). The haphazard collection of nineteenth- and early twentieth-century American fiction and *belles lettres* has been included mainly to show that the spread of gerunds is a recent phenomenon even in American English – the more so as many of the older works are narratives and would thus have provided the past-tense environment now favouring the gerund.

Again it is helpful to interpret the complexities of current shifting usage against the long-term history of English aspectual verbs like *begin* and *start*. Thereby, *start* is the simple case, because its use as an aspectual verb is recent (eighteenth century), and it seems to have freely combined both with infinitives and gerunds since that time. Similarly uncontroversial is the history of the *begin* + infinitive-construction. As Brinton (1988: 109–11) has shown, the construction has been widely attested since Old English times. Up to and including the Middle English period, *begin* could take both the bare infinitive and the infinitive with *to* or (in Middle English) *for to* (Brinton, 1988: 118–19). Since then, the infinitive with *to* has been the norm.

The history of gerundial complements with *begin*, by contrast, is somewhat obscure. Visser suggests a Middle English origin of the construction (1970–8, III: 1888–90), adduces some examples going back to the period between ca. 1205 to 1553, notes a gap of two centuries in the record and then provides continuous documentation from 1813 onwards. However, he voices doubts about some of his early examples himself, and I think that a case could well be made that all of them are irrelevant, including the two explicitly singled out as good by Brinton (1988: 119). They all probably represent either nominal *-ing*-phrases (cf., e.g., modern *I began square-dancing*) or *-ing*-phrases that are adverbial clauses (cf. modern *I began immediately, saying . . .*).

The alternative is to see *begin* + V-*ing* as a late eighteenth- or early nineteenth-century innovation. This view is consistent with the evidence provided by Jespersen's *Modern English grammar*, whose earliest examples of *begin* + V-*ing* are from nineteenth-century British English (IV: 171, V: 196), the relevant *OED* entry, which does not attest any gerundial complements, and, finally, the fact that *begin* + V-*ing* is not attested at all in the Early Modern English component of the Helsinki Corpus.[9]

Surveying verbs of inception and a number of other semantic classes, Fischer (1995) has suggested that modern V-*ing* complements are functional continuations of the now obsolete bare infinitival complements. This tempting functionalist

are only three relevant uses of *begin* (2 gerunds, one infinitive), while the more frequent *start* shows the expected distribution: 17 gerundial complements against only 3 infinitival ones, which moreover involve precisely those syntactic environments most strongly disfavouring the gerund (*start* being used as a participle itself, inanimate subjects).

[9] Teresa Fanego, personal communication. She also points out that in spite of the evidence on *begin* the use of gerunds with verbs of inception has a tenuous foothold even in Early Modern English, for example in the relevant use of *fall*, as in *fall a-laughing* or *fall to laughing*. I agree that, through providing analogical support, this construction may have been a minor factor in the establishment of *begin* + V-*ing*, but the similarities should not be over-emphasized, as with *fall* a preposition or at least a residual proclitic *a-* before the participle is obligatory.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

explanation of the rise of V-*ing* complements as a compensatory strategy is compatible only with Visser's assumption of an early origin of *begin* + V-*ing*. Therefore, the choice between the two rival historical accounts has some theoretical repercussions, and it is important to shed light on the crucial two and a half centuries (1553–1813) missing from Visser's record. A search for relevant citations in the quotation base of the electronic edition of the *OED* yields the following instructive results.

Begin + V-*ing* is infrequent even in the modern material from the nineteenth and twentieth centuries, and extremely rare before that time. A total of 2,747 citations containing *begin* and 1,575 containing *begins*[10] yielded no more than 76 gerund complements: 6 pre-1800, 30 for the nineteenth century, and 40 for the twentieth. As the nineteenth century is better represented in the quotation base than the twentieth (763,987 against 481,376 quotations), these figures are evidence of a fairly quick spread of the gerund in the course of the past century, a trend which in our four corpora was apparent in the American material.[11]

Of the 6 pre-1800 examples, 5 are from the period between 1771 and 1794 and would thus seem to support a late origin rather than continuity from Middle English. The one potential missing link to Visser's alleged Middle English examples is from the early seventeenth century:

(6) 1630 J. Taylor (Water P.) Bk. Martyrs Wks. iii. 141/1, So our Eliza stoutly did begin Untopping and beheading Romish sin.

However, this example needs to be classed with the dubious ones. An infinitive would have ruined the metre of the couplet, and an analysis of the *-ing*-phrase as adverbial is plausible. We are not dealing with a clear instance of an aspectual construction. So as not to overlook pertinent evidence, I undertook a further search in the OED quotation base on *began* and *begun*, limiting myself this time to the crucial period from 1600 to 1800 for obvious reasons of expediency. Of the 12 plausible gerundial complements found, 9 were from the second half of the eighteenth century (post-1743, to be precise). One, from Fuller's 1655 *Church-history of Britain,* is uninterpretable for the purposes of the present argument because of cuts made by the *OED*'s editors and the unavailability of a copy of Fuller's original at the time of this writing, and the following two remain as potential missing links:

(7) 1614 SYLVESTER Du Bartas, *Bethalia's rescue* iii.1 *Flame-snorting Phlegon's ruddy breath began Reducing Day.
(8) 1711 Swift, *Jrnl Stella* 7 July, It began raining, and I struck into Mrs. Vanhomrigh's, and dined.

[10] Since the figures reported below speak for themselves, I decided not to sift through the several thousand remaining occurrences of *began*, *beginning* and *begun*, or their many orthographic variants, although – as was shown (cf. table 6) – the form of the verb has some impact on whether it is followed by an infinitive or gerund in Modern English.

[11] As it is extremely laborious to assign citations to their geographical origins in the *OED* material, I have not attempted to assess to what extent the observed overall increase is due to British or American influence.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Example (7), from 1614, is from a translation of a poetic text and thus not beyond doubt. With its impersonal verb, example (8), from 1711, however, is a perfect instance of the use of *begin* as an aspectual semi-auxiliary and – to my knowledge – represents the earliest uncontroversial instance. In any event, a mere two (or even three[12]) good examples are not enough to forge a link of continuity between Visser's Middle English examples and the later eighteenth century, when the *begin* + V-*ing* construction begins being attested continuously again.

Establishing a late origin for *begin* + V-*ing* thus means that, at least for this particular verb, we must dismiss Fischer's hypothesis that the gerund complement replaces a bare infinitival one. In addition, the present case now reveals obvious parallels to the one studied in the preceding section. It seems that, just like *prevent* + NP + V-*ing*, *begin* + V-*ing* arose at the time when the future British and American standards of English were differentiating from a common base. Unlike phonetics, where the differentiation of the British and American standards of pronunciation was already well under way, the grammatical identities of the two varieties are more recent. They have gradually emerged only in the course of the past two centuries and involve developments which in some instances have not run to completion yet.

6 Americanization *cum* grammaticalization: *help* + bare infinitive

In terms of the typology of possible results presented above, the distribution of bare and *to*-infinitival complements after *help* represents a mixed type. What we have in the short term is an approximation of current British usage to US norms (that is constellation 4, 'Americanization'); in the long term, however, this Americanization is embedded in an overarching 'parallel diachronic development' leading to an increase in frequency of bare infinitives in all varieties (constellation 3).

In 1961, the publication date of the texts assembled in the Brown and LOB corpora, the *to*-infinitive was the statistical norm in British English, whereas the bare infinitive dominated in American English (Algeo, 1988: 22; Kjellmer, 1985; table 7 below). This is also the state of things recorded in Quirk et al.'s widely used reference grammar. Commenting on the variation between *Sarah helped us edit the script* and *Sarah helped us to edit the script*, the authors say:

> Of the two constructions with *help*, that with *to* is more common in BrE, and that without *to* is more common in AmE. (Quirk et al., 1985: 1205f.)

[12] Teresa Fanego points out the following additional example from Louis T. Milic's Century of Prose corpus:

> On Monday next the lottery will begin drawing, and fortune to dispense her golden showers on those that have the prudence to hold out their dish. (*Daily Advertiser*, 1741)

While the form *drawing* itself is not unambiguously verbal (after all, there is no following object or modifying adverbial), the mediopassive semantics of the verb *draw* does suggest such an interpretation. Note also the sequence of coordinated verbal complements in which the first element is realized as V-*ing* and the second is an infinitive. Structures of this type are commonly encountered even today, whereas the converse – infinitive followed by a gerund – is extremely rare.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table 7. to- *vs. bare infinitives in four corpora*

	BrE	AmE
1961	94:27	55:125
1991/92	77:122	44:203

(BrE vs. AmE 1961 $p<0.001$; BrE vs. AmE 1991/92 $p<0.05$; BrE diachr. $p<0.001$; AmE diachr. $p<0.001$)

Table 8. *Complementation of* help *in the 'spoken-demographic' BNC*

	without following NP/object	with following NP/object	total
help + bare infinitive	34	92	126
help + *to*-infinitive	22	44	66

The results from our corpora show that while this may have been true in the 1960s, it is no longer so now.[13] The bare infinitive is now the statistical norm also in British English.

This result is not an artefact of the corpus consulted. The bare infinitive was found to be the more frequent form in other recent text databases containing written British English (Mair, 1995: 268, on the *Guardian* on CD-ROM), and it is the more common form also in 1990s spoken British English – as evidenced by the figures in table 8 from the spoken-demographic sample of the British National Corpus, which contains roughly 4 million words of orthographically transcribed spontaneous speech.

In the British tradition, where the distinction is not always rigidly made between American, informal or uneducated usage,[14] it has been commonplace to consider the

[13] Since I argued above that participant observers unaided by corpora are in a weak position when monitoring ongoing grammatical change, fairness requires me to draw attention to the following very perceptive analysis in Foster (1968: 204): 'the constructions accompanying certain verbs quietly change over the years without causing any great outcry. Some notable changes of this sort are once again products of American idiom, a typical example being seen in the omission of the preposition "to" after "help". Now this phenomenon was not unknown in poetical and somewhat archaic language . . . . But only in the late nineteen-thirties and early 'forties did the construction really make headway in Britain. Its acceptance into the standard language was very rapid and J. Hubert Jagger, writing his *English in the Future* (1940), commented on "the speed with which the American habit of omitting *to* after *help* has invaded Britain" (p. 55). But in spite of the speedy acceptance of the new form the old one is still well entrenched and the two rivals seem destined to battle it out for some time to come.' The 1930s and 1940s were indeed important in the spread of the bare-infinitive construction in British English (cf. the frequency distributions in the OED discussed below), and one wonders about the role of the common contemporary collocation *help us (to) win the war* in the process.

[14] As one source graciously puts it: '*Help* followed by an infinitive without *to* . . ., once condemned as an Americanism, is now accepted in British English . . .' (Wood, 1962: 107).

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

bare infinitive as in some way informal or nonstandard, as opposed to the formal or standard construction with *to*. Particularly instructive in this connection are changes to the relevant *OED* entry, where in the first edition of 1933 the bare infinitive figures as dialectal and obsolete, whereas the new edition of 1989 lists it as a common colloquial form.[15] In view of the parallels between current written and spoken usage in Britain an analysis of the variation between the two constructions as stylistic seems difficult to maintain.

As in the case of *begin* and *start*, there is a rich literature in which various structural, semantic or iconic factors are discussed with regard to their influence on the synchronic distribution of bare and *to*-infinitives after *help* (cf., e.g., Dixon, 1991: 199, 230; or Duffley, 1992: 29). There is not enough space here to engage with the arguments in detail, but it cannot come as a surprise that any such synchronic account must remain limited when we are faced with an obvious instance of change in progress.

All analyses proposed so far – including those that correctly diagnose the disappearance of a regional contrast between British and American English – have disregarded a second important fact. It is not only British English which has been changing by moving closer to American English; American English itself has been developing, as well. Infinitival complements after *help* do not form a closed system in which the proportion of bare and *to*-infinitives may change but the over-all frequency of the relevant instances remains constant. Even in the course of the very short period documented in the four corpora studied here, instances of *help* governing (any kind of) infinitive have increased significantly – from 121 to 199 in the British corpora, and from 180 to 247 in the American ones (see table 7).

This increase is not a statistical fluke but part of a long-term trend which emerges very clearly from the quotation base of the *OED* (see figure 2 below). To compensate for the fact that the number of quotations per period is not constant, frequencies are normalized, giving instances per 10,000 quotes (see Appendix for full figures). Barring some fluctuations in the proportions of bare and *to*-infinitives, nothing happens for the first two and a half centuries of the period under review. Instances of *help* + infinitive never exceed a very low frequency of 5 per 10,000 citations. From the mid nineteenth century onwards, however, uses of *help* with infinitival complements start mushrooming.[16] Overall, increase is fastest for the bare infinitives.

The higher frequency of *help* is not due to the fact that it has ousted synonyms such as *support* or *aid*. Nor can the infinitives be said to have encroached on *that*-clauses, as these are rare in Old and Middle English and absent from the Early Modern period onwards. The only plausible explanation is that the meaning of *help* has broadened, from 'somebody lends support to somebody else in performing some

[15] As will be shown below, both assessments are in stark contrast to the evidence provided by the dictionary's own quotation base at the time they were made.

[16] As the twentieth century is plotted by decades rather than 25-year intervals, the 'real' rise is even sharper.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

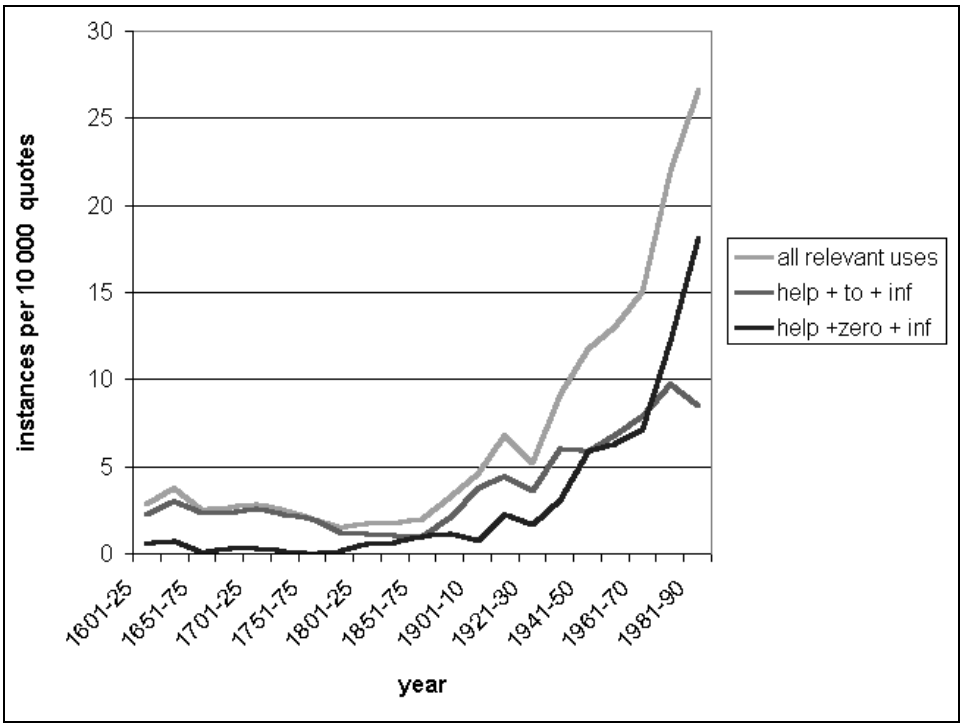


Figure 2. *help* + infinitival complements in the *OED*

task' to a more general notion of 'contribute to/provide a favourable environment for'. It will be noted that while the first meaning is compatible with inanimate subjects or objects only in metaphorical diction, such constraints are absent for the second. In fact, the latter meaning is so general and abstract that it approaches those typically associated with grammatical categories. The verb *help* might thus be said to be in the process of taking over quasi-auxiliary function in complex verb phrases (see Mair, 1995 for a tentative proposal of this kind).

Reviewing research by Benveniste and others, Brinton points out that the creation of new auxiliaries by grammaticalization crucially involves three kinds of re-analysis:

> (a) of a full verb as an auxiliary,
> (b) of a participle or infinitive as an 'auxiliate', and
> (c) of a loose concatenation of main verb plus verbal complement as a unified or frozen form. (Brinton, 1988: 96f.)

Given the grammatical facts of English, in which true auxiliaries are a closed class of anomalous finites with a large number of clear morphosyntactic properties, avenue (c) is the pertinent one here. The *help* + verb combination is about to be added to the large number of modal idioms and catenatives already in existence in Modern English. The fact that the reduced form of the infinitive (typically, but not

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

exclusively, found with modals in Modern English) has become the statistical norm is a telling sign that the process is already well under way.

Brinton's avenue (b) is relevant indirectly. It has long been noted that the bare infinitive is more likely to be used when the verb *help* itself is in the infinitive, because in this way a sequence of two *to*-infinitives may be avoided (cf., e.g., Rohdenburg, 1995a: 380–2, with further references to the literature). As a stand-alone explanatory device, this *horror aequi* constraint is not entirely convincing, because it is not at all difficult to find examples of *to help to*. In the British National Corpus (100 million words), there are 132 such cases, for example, of which only very few are dubious or spurious. It is reasonable to analyse most such *to help to+inf-* sequences as 'auxiliates' in Brinton's sense, and not as two separate infinitival clauses arranged in sequence, and it is clear that in such cases considerations of euphony or a desire to avoid processing problems – the two explanations usually given for strong *horror aequi*-effects – will not play a major role.

If, on the short-term analysis, our diagnosis was that a contrast separating British and American usage until well into the 1960s was levelled by the 1990s (convergence on the American norm, constellation 4), this view needs to be modified now. Both British and American English are developing along the same lines (parallel diachronic development, constellation 3), and the contrast was a transitional phenomenon, due to the fact that the development proceeded at different speeds in the two varieties for a time.

By way of conclusion, a few typical examples will be discussed of the kind of use which is responsible for the increase in the frequency of *help*. They are taken from the quotation base of the *OED*:

(9) 1941 Punch 2 July 13/3 Sir Kingsley Wood . . . asked the House for another £1,000,000,000, to help pay for the next three months of war.

(10) 1961 L. Mumford City in History xv. 479 Nor have they eliminated the unburned hydrocarbons which help produce the smog that blankets such a motor-ridden conurbation as Los Angeles.

(11) 1968 National Observer (US) 8 Apr. 5/4 Negro cabbie John W. Smith, whose arrest for 'tailgating' a police car . . . helped spark five days of rioting . . ., was found guilty of assaulting a policeman.

(12) 1976 Alyn & Deeside Observer 10 Dec. 5/2 Part of the fun of the game comes in 'sooping'. This is when the players sweep the ice with special brooms in front of a moving stone to help it go further.

Example (9) illustrates the pseudo-prepositional use of the infinitive. '[Money] *to help pay for* the next three months of war' is '[money] *for paying* for the next three months of war'. Formulated as it is, the sentence suggests a structure in which an instrument has been promoted to the syntactic role of subject: it is the money that pays for the war, and we do not think about the actual agent who spends the money in order to pay for the war. Inserting *to* before *pay* in this example would not only be stylistically clumsy because of the repetition involved; it would also produce a slight shift in perspective, from the instrument (money) to the agent who spends it. In 'Sir Kingsley Wood asked the House for another £1,000,000,000, *to help to pay*

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

for the next three months of war', the relevant semantic frame for the interpretation of *help* is more likely to be that associated with the literal three-place predicate: by granting the money, the House helps Sir Kingsley/the Government to pay for the war. Examples (10) and (11) feature negative effects – smog and rioting – which are not compatible with the core semantics of *help*: nobody is helped/supported in order to produce smog or spark off a riot here, which is why adding *to* before the infinitives would be slightly incongruous. Rather, inanimate entities create a favourable environment for the negative effects. Example (12), finally, is a fairly clear case of a purely causative use of *help*, equivalent to *make* ('make it go further'). Again, adding *to* before the infinitive is problematical.[17]

7 Conclusion

The aim of the present article was to show that a set of matching corpora, together with further machine-readable databases, is a useful resource in the study of ongoing grammatical change in standard English. For all three phenomena subjected to detailed study, it was possible to add to, or to correct, existing descriptions. This was due to the fact that the corpora provided large and diachronically layered amounts of authentic data which could be analysed quantitatively and, where necessary, qualitatively. The rich corpus-linguistic working environment available to the student of standard English has made possible an integrated description of the synchronic and diachronic factors at work in the observed variation.

A few generalizations going beyond the individual features and variables studied are possible. It seems that, in contrast to phonetics, where the record suggests an independent American norm already for the eighteenth century, grammatical contrasts between the two varieties are more recent. They are usually statistical tendencies rather than firm rules, and in several cases still emerging. Where received opinion would have it that World English is converging on American norms, the detailed view reveals a complex dialectic of convergence (in our case, the use of *help* + infinitive) and divergence (e.g. *prevent*).

Corpus evidence is particularly useful for the study of ongoing processes of grammaticalization. In this respect, the present results can be placed alongside related work on ongoing change by Olofsson (1990), Rickford et al. (1995) or Romaine & Lange (1991). In this way, evidence builds up to prove that suspected grammaticalization processes can in fact be empirically verified while they are progressing – a claim explicitly rejected by several theorists (cf., e.g., Lehmann, 1991: 532; or Compes, Kutscher & Rudorf, 1993: 20).

[17] One anonymous referee correctly points out that the semantic contrasts associated with the presence or absence of *to* in the above examples represent a vindication of 'semantic' approaches to English complementation such as Dixon (1991) and Duffley (1992), which have been criticized above. I take the point, but I would still maintain that against a background of change in progress the contrasts in question will usually be too unstable and unsystematic to provide an exclusive basis for a grammatical description.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Ultimately, the study of ongoing grammatical change in well-documented languages yields the type of 'thick description' (Clifford Geertz) that provides the ideal testbed for utterance-based models of linguistic evolution such as those recently proposed by Keller (1990) or Croft (2000). In this way, detailed insights into language-specific developments such as the ones presented here can feed back into the theoretical debate on language change.

Author's address:
Christian Mair
Albert-Ludwigs-Universität Freiburg
Englisches Seminar 1
D – 79085 Freiburg
Germany
christian.mair@anglistik.uni-freiburg.de

APPENDIX

Table A. *OED*/help *(twentieth century): absolute frequencies*

decade	all relevant uses	*to* total	+NP	−NP	zero 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

Table B. *OED*/help *(twentieth century): proportional frequencies (= instances per 10,000 quotes)*

decade	number of quotations	all rel. uses	*to* total	+NP	−NP	zero 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

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table C. *OED*/help *(nineteenth century): absolute frequencies*

period	all relevant uses	*to* total	+NP	−NP	zero 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

Table D. *OED*/help *(nineteenth century): proportional frequencies (= instances per 10,000 quotes)*

period	number of quotations	all rel. uses	*to* total	+NP	−NP	zero 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		

Table E. *OED*/help *(eighteenth century): absolute frequencies*

period	all relevant uses	*to* total	+NP	−NP	zero 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	–

Table F. *OED*/help *(eighteenth century): proportional frequencies (= instances per 10,000 quotes)*

period	number of quotations	all rel. uses	*to* total	+NP	−NP	zero 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		

Table G. *OED*/help *(seventeenth century): absolute frequencies*

period	all relevant uses	*to* total	+NP	−NP	zero 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

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Table H. *OED*/help *(seventeenth century): proportional frequencies (= instances per 10,000 quotes)*

period	number of quotations	all rel. uses	*to* total	+NP	−NP	zero 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		

References

Algeo, J. (1988). British and American grammatical differences. *International Journal of Lexicography* **1**: 1–31.

Algeo, J. (1990). The Briticisms are coming! How British English is creeping into the American language. *Journal of English Linguistics* **23**: 123–40.

Altenberg, B. (1991). A bibliography of publications relating to English computer corpora. In Johansson, S. & A.-B. Stenström (eds.), *English computer corpora: selected papers and research guide*. Berlin: Mouton de Gruyter. 355–96.

Bailey, C. N. (1979). Developments in British English grammar. *Arbeitsberichte zur Sprachentwicklung und Sprachbeschreibung / Technische Universität Berlin Working Papers in Linguistics* **4**: 133.

Barber, C. (1964). *Linguistic change in present-day English*. London and Edinburgh: Oliver & Boyd.

Barlow, M. (1998). *A corpus of spoken professional American English*. Houston, TX: Athelstan.

Bauer, L. (1994). *Watching English change: an introduction to the study of linguistic change in standard Englishes in the twentieth century*. London: Longman.

Biber, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

Biber, D. & E. Finegan (1989). Drift and evolution of English style: a history of three genres. *Language* **65**: 487–517.

Biber, D., S. Conrad & R. Reppen (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D. et al. (1999). *The Longman grammar of spoken and written English.* Harlow: Longman.

Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart & Winston.

Brinton, L. (1988). *The development of English aspectual systems: aspectualizers and postverbal particles*. Cambridge: Cambridge University Press.

Compes, I., S. Kutscher & C. Rudorf (1993). *Pfade der Grammatikalisierung: ein systematischer Überblick*. Arbeitspapier 17. Köln: Institut für Sprachwissenschaft der Universität.

Croft, W. (2000). *Explaining language change: an evolutionary approach*. London: Longman.

Dixon, R. M. W. (1991). *A new English grammar, on semantic principles*. Oxford: Oxford University Press.

Dixon, R. M. W. (1995). Complement clauses and complementation strategies. In Palmer, F. R. (ed.) *Grammar and meaning: essays in honour of John Lyons.* Cambridge: Cambridge University Press. 175–220.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

Duffley, P. J. (1992). *The English infinitive*. London: Longman.
Duffley, P. J. (1999). The use of the infinitive and the *-ing* after verbs denoting the beginning, middle and end of an event. *Folia Linguistica* **33**: 295–331.
Fanego, T. (1997). On patterns of complementation with verbs of effort. *English Studies* **78**: 60–7.
Fanego, T. (1998). Developments in argument linking in early Modern English gerund phrases. *English Language and Linguistics* **2**: 87–119.
Fischer, O. (1995). The distinction between *to* and bare infinitival complements in late Middle English. *Diachronica* **12**: 1–30.
Francis, W. N. & H. Kučera (1982). *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin.
Foster, B. (1968). *The changing English language*. London: Macmillan.
Greenbaum, S. (1986). The *Grammar of contemporary English* and *A comprehensive grammar of the English language*. In Leitner, G. (ed.), *The English reference grammar*. Tübingen: Niemeyer. 6–14.
Gronemeyer, C. (1999). On deriving complex polysemy: the grammaticalization of *get*. *English Language and Linguistics* **3**: 1–39.
Hansen, K. et al. (1996). *Die Differenzierung des Englischen in nationale Varianten: eine Einführung*. Berlin: Schmidt.
Hundt, M. (1997). Has British English been catching up with American English over the past thirty years? In Ljung, M. (ed.), *Corpus-based studies in English: papers from the 17th International Conference on English Language Research Based on Computer Corpora*. Amsterdam: Rodopi. 135–51.
Hundt, M. & C. Mair (2000). 'Agile' and 'uptight' genres: the corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* **4**: 221–42.
Jespersen, O. (1909–49). *A modern English grammar on historical principles*, 7 vols. Copenhagen: Munksgaard.
Keller, R. (1990). *Sprachwandel: von der Unsichtbaren Hand in der Sprache*. Tübingen: Francke.
Kirsten, H. (1957). Bemerkungen zu *to prevent* + Gerundium. *Zeitschrift für Anglistik und Amerikanistik* **1**: 327–30.
Kjellmer, G. (1985). *Help to/help Ø* revisited. *English Studies* **66**: 156–61.
Labov, W. (1981). What can be learned about change in progress from synchronic description? In Sankoff, D. & H. Cedergren (eds.), *Variation omnibus*. Edmonton, Alberta: Linguistic Research. 177–201.
Labov, W. (1994). *Principles of linguistic change*, vol. 1: *Internal factors*. Oxford: Blackwell.
Lehmann, C. (1991). Grammaticalization and related changes in contemporary German. In Traugott, E. C. & B. Heine (eds.), *Approaches to grammaticalization*, vol. 2. Amsterdam: Benjamins. 493–535.
Lightfoot, D. (1991). *How to set parameters: arguments from language change*. Cambridge, MA: MIT Press.
Mair, C. (1995). Changing patterns of complementation, and concomitant grammaticalisation, of the verb *help* in present-day British English. In Aarts, B. & C. F. Meyer (eds.), *The verb in contemporary English: theory and description*. Cambridge: Cambridge University Press. 258–72.
Mair, C. (1997). Parallel corpora: a real-time approach to the study of language change in progress. In Ljung, M. (ed.), *Corpus-based studies in English: papers from the 17th International Conference on English Language Research Based on Computer Corpora*. Amsterdam: Rodopi. 195–209.
Mair, C. & M. Hundt (1995). Why is the progressive becoming more frequent in English? A

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press

corpus-based investigation of language change in progress. *Zeitschrift für Anglistik und Amerikanistik* **43**: 111–22.
Marckwardt, A. H. (1980). *American English*, 2nd edn, rev. J. L. Dillard. New York: Oxford University Press.
McEnery, T. & A. Wilson (2001). *Corpus linguistics*, 2nd edn. Edinburgh: Edinburgh University Press.
Mindt, D. (2000). *An empirical grammar of the English verb*. Berlin: Cornelsen.
Olofsson, A. (1990). A participle caught in the act: on the prepositional use of *following*. *Studia Neophilologica* **62**: 23–35, 129–49.
Potter, S. (1975). *Changing English*, 2nd edn. London: Deutsch.
Pullum, G. & A. Zwicky (1999). Gerund participles and head-complement inflection conditions. In Collins, P. & D. Lee (eds.), *The clause in English: in honour of Rodney Huddleston*. Amsterdam: Benjamins. 251–71.
Quirk, R. (1974). *The linguist and the English language*. London: Arnold.
Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985). *A comprehensive grammar of the English language*. London: Longman.
Rericha, V. (1987). Notes on infinitival and *-ing* complements of the verbs *begin* and *start*. *Philologica Pragensia* **30**: 129–32.
Rickford, J., N. Mendoza-Denton, T. Wasow & J. Espinosa (1995). Syntactic variation and change in progress: loss of the verbal coda in topic restricting *as far as* constructions. *Language* **71**: 102–31.
Rohdenburg, G. (1995a). On the replacement of finite complement clauses by infinitives in English. *English Studies* **76**: 367–88.
Rohdenburg, G. (1995b). Bemerkungen zum Auf- und Abstieg einiger präpositionaler Konstruktionen des Englischen. *NOWELE: North Western European Language Evolution* **26**: 67–124.
Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* **7**, 149–82.
Romaine, S. & D. Lange (1991). The use of *like* as a marker of reported speech and thought: a case of grammaticalization in progress. *American Speech* **66**: 227–79.
Schmid, H.-J. (1993). Cottage *und Co*., idea, start *vs.* begin: *Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung*. Tübingen: Niemeyer.
Strang, B. (1970). *A history of English*. London: Routledge.
Taylor, L., G. Leech & S. Fligelstone (1991). A survey of English machine-readable corpora. In Johansson, S. & A.-B. Stenström (eds.), *English computer corpora: selected papers and research guide*. Berlin: Mouton de Gruyter. 319–54.
Trudgill, P. & J. Hannah (1982). *International English*, 2nd edn. London: Edward Arnold.
Visser, F. Th. (1970–78). *An historical syntax of the English language*, 3 vols. Leiden: Brill.
Webster's dictionary of English usage. 1989. Springfield, MA: Merriam-Webster.
Wilson, K. G. (1993). *The Columbia guide to English usage*. New York: Columbia University Press.
Wood, F. T. (1962). *Current English usage*. London: Macmillan.

https://doi.org/10.1017/S1360674302001065 Published online by Cambridge University Press