

Change in the English infinitival perfect construction

Jill Bowie and Bas Aarts
UCL

1 Introduction

The availability of searchable electronic corpora composed of textual material from different periods in time has made studying change in the English language easier. However, as is reflected in chapters of this volume, there are a number of methodological dimensions to the use of corpora in the study of current change. For some 'big is better' (Davies, this volume), while for others 'small is beautiful' (Hundt and Leech, this volume; Smith and Leech, forthcoming). Our own position is that, while we can see the distinct advantages of using large corpora, detailed analysis of small corpora, especially if they are parsed, can reveal trends that may be missed by other approaches. Furthermore, in studying changes over short periods we believe that spoken language corpora are particularly valuable, as spoken language is primary, and changes in grammar are likely to manifest themselves in that medium first.

This paper explores short-term changes in the English infinitival perfect construction. It first examines changing frequencies of occurrence, comparing the various tense forms (present, past and non-finite) of the perfect. It then focuses on the infinitival perfect, considering its syntactic contexts of occurrence and qualitative explanations for the changes observed. The data are drawn from the *Diachronic Corpus of Present-day Spoken English* (DCPSE), based at the Survey of English Usage, University College London. DCPSE is unique in several ways. First, it contains exclusively spoken (and mainly spontaneous) English, in two subcorpora with matching text categories which allow diachronic comparison over a thirty-year span. The earlier subcorpus contains approximately 464,000 words from the *London-Lund Corpus* (LLC) dating from the late 1950s to the 1970s, while the later subcorpus contains around 421,000 words from the *British Component of the International Corpus of English* (ICE-GB) collected in the early 1990s. Secondly, the corpus is fully parsed and searchable with dedicated corpus exploration software called ICECUP (International Corpus of English Corpus Utility Program).¹

2 The perfect construction in DCPSE

The English perfect construction involves the perfect auxiliary HAVE followed by a verb in the past participle form. It occurs in present, past and non-finite forms, all of which typically function to express anteriority (i.e. pastness relative to a reference point). The present perfect (*I **have read** the book*) generally presents a situation as occurring within (or even continuing through) a timespan beginning in the past and leading up to the present, with an added dimension of 'current relevance' (i.e. a focus on the present repercussions of the situation). The past perfect (*By the time he returned I **had read** the book*) typically encodes pastness relative to a reference point in the past, and the non-finite perfect indicates anteriority in various types of construction (*He must **have read** the book; **Having said** that, I still like her*). The present perfect has developed a specialised use through contrast with the (morphologically marked) past tense (used to present a time as wholly in the past rather than connected to the present). This contrast, however, is neutralised in the non-present perfects, which can correspond to either a present perfect or a simple past (Huddleston and Pullum et al. 2002: 146). This is illustrated by the following examples of the infinitival perfect, where (1) corresponds to a present perfect (with

¹ See Svartvik (1990) on LLC and Nelson et al. (2002) on ICE-GB and ICECUP; for more information see www.ucl.ac.uk/english-usage.

since the election specifying a period up to the speaker's present) and (2) to a simple past (where a specific time in the past is under discussion):

- (1) tick off for me the main things that you would claim to **have achieved** as a government since the election [DL-E02#0027]²
- (2) and Mr Perry [...] claimed [...] to **have heard** a great deal of noise from this motor-cycle as it came along followed by the bang [DL-J04#0099]

Examples of the perfect construction can be retrieved in DCPSE by using Fuzzy Tree Fragments (FTFs), a search facility within the ICECUP software (Aarts et al. 1998; Nelson et al. 2002). This facility allows the user to construct partial tree diagrams and to choose the level of detail specified (hence 'fuzzy'). Perfect auxiliaries can be found through a simple FTF search for a single node of category 'AUX' with the type feature 'perf'. Different tense features can then be added to this node to search more specifically for instances occurring in present tense, past tense, or one of the non-finite forms. Figure 1 shows an FTF used to search for the infinitival form. Categorial information (such as word or phrase class) is shown in the upper righthand box of an FTF, functional information (such as direct object, noun phrase head) in the top lefthand box, and additional features in the lower box. In this instance the function has been left unspecified.

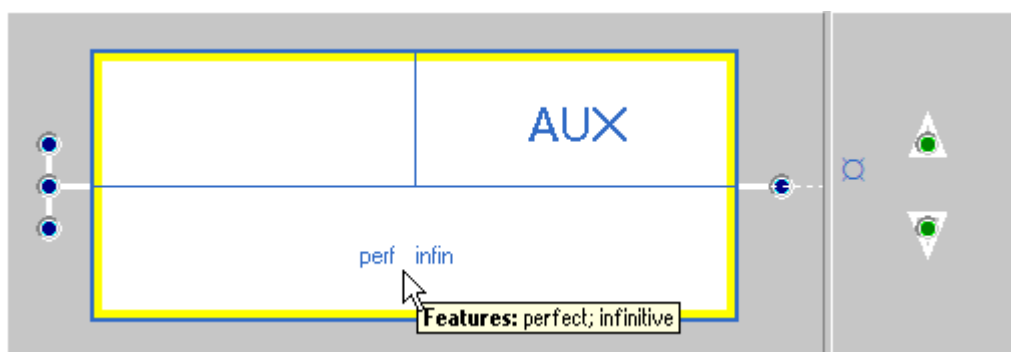


Figure 1: A simple FTF to search for perfect infinitive auxiliaries.

Frequencies (normalised per million words, 'pmw') can then be compared for LLC (the earlier subcorpus) and ICE-GB (the later subcorpus). The results show that the perfect auxiliary falls in frequency by nearly 8% across the two subcorpora (Table 1, Total row).³

Tense category	LLC		ICE-GB		Change in frequency		
	raw	pmw	raw	pmw	%	A: χ^2 (words)	B: χ^2 (perfect)
present	3,572	8,020.17	3,343	8,277.17	+3.20%	1.72 ns	21.71 s
past	835	1,874.82	484	1,198.37	-36.08%	62.40 s	41.74 s
infinitive	652	1,463.93	413	1,022.58	-30.15%	32.90 s	19.71 s
-ing participle	78	175.13	58	143.61	-18.00%	1.31 ns	0.46 ns
Total	5,137	11,534.05	4,298	10,641.72	-7.74%	15.18 s	

Table 1. Frequencies of perfect auxiliaries in DCPSE, divided by tense category. Columns A and B represent goodness of fit χ^2 comparisons summarised in the text. Results marked 's' are significant at $p < 0.05$. Figures include occurrence of HAVE in the stative idiom HAVE *got* (see discussion in text).

² Examples from DCPSE are cited with their identifying text codes (prefixes 'DL' and 'DI' indicate LLC and ICE-GB subcorpora respectively).

³ Note that the data reported in this paper are based on a revised version of DCPSE prepared by the authors and others at the Survey of English Usage.

However, when the data for the different tense forms are considered, an unexpected picture emerges. What Table 1 demonstrates is that not every tense form behaves in the same manner. Past and infinitival forms fall by around 36% and 30% respectively, whereas the present (by far the most frequent of the forms) is stable. We carry out two distinct series of chi-square tests: in Column A we compare the distribution of each term (present, past, etc.) with the total number of words; in Column B we compare each term relative to the trend of the overall set of perfect auxiliaries.⁴ Note that the slight percentage increase (3.20%) of the present tense is not considered significant compared with the number of words (Column A), but *it does differ from the overall pattern* (Column B). The figures for the *-ing* participle are small and neither result is significant. Figure 2 displays the changes in pmw frequencies, with confidence intervals depicted by error bars.⁵

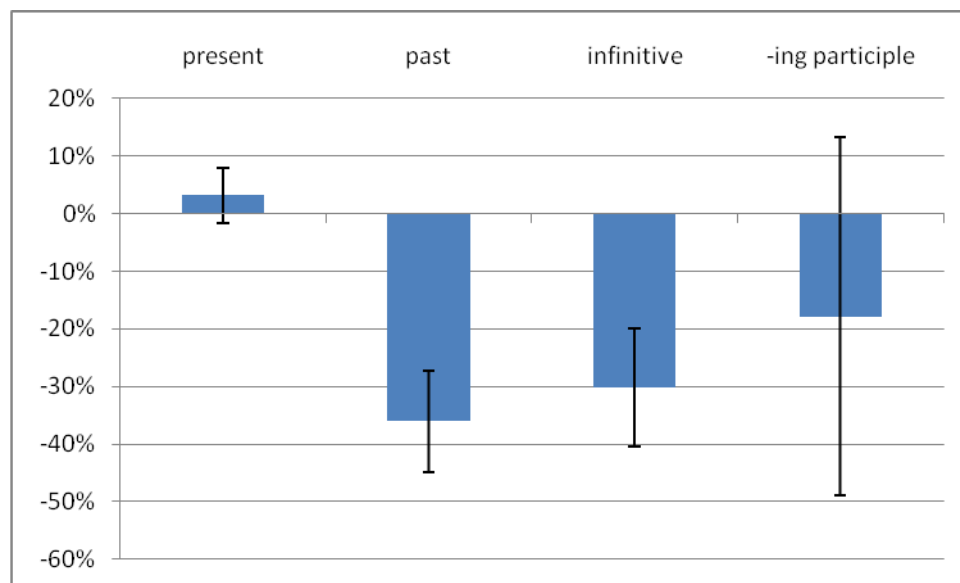


Figure 2: Changes in pmw frequencies (Table 1 ‘%’ column) with error bars for $p < 0.05$. Where an error bar does not cross the zero axis, the change is statistically significant (cf. Table 1 Column A).⁶

In considering these data, the frequent combination HAVE + *got* merits attention. Instances of the ‘semi-modal’ HAVE *got* [to] (as in *a lot of work **has got to be done on it***) are automatically excluded by the FTF searches, as it is analysed in the corpus as a semi-auxiliary (an auxiliary with the type feature ‘semi’). However, the figures do include, alongside clear instances of the perfect construction (e.g. *How advanced **have they got***), instances where the combination takes an NP object and expresses a stative meaning, such as *he’s **got two kids*** (‘he has two kids’). This use is described by Quirk et al. (1985: 131) as ‘perfective in form’ but ‘nonperfective in meaning’, and by Huddleston and Pullum et al. (2002: 111–13) as an idiom which is only historically a perfect construction. We therefore used FTFs to find instances where HAVE, parsed as a perfect auxiliary, is followed by *got* (allowing for intervening material such as adverbs). These are very frequent in the present tense category of our perfect data set, comprising nearly 23% of examples (with a slightly higher proportion in ICE-GB than in LLC), and examination

⁴ To be more precise, in Column A we carry out a goodness-of-fit χ^2 test (Sheskin 1997: 95) for the overall change (in the **Total** row) and for each individual subcategory (present, past, infinitive and *-ing* participle) against the number of words in the corpus. This evaluates whether the observed percentage change is significant (i.e. significantly *different from zero*; see Figure 2). Column B uses the same test against a perfect auxiliary baseline.

⁵ These were computed using the Newcombe (1998) proposed interval for the difference between two proportions, which is based on the Wilson score interval. This is a more precise method than traditional Gaussian error bars.

⁶ This type of visualisation displays the size of the result (i.e. the column height) and our confidence in it. Another way of expressing this is that we are 95% certain that, for example, the *past perfect* case falls by between –27% and –45%.

of a 10% random sample showed that a majority are stative or ambiguous (cf. Biber et al. 1999: 463–7, who include stative examples in their present perfect counts, and note their high frequency in British English conversation). Occurrences with *got* comprise less than 7% of our data for each of the other tense form categories; some of the past tense examples are stative or ambiguous, but there are no clear stative examples (and only one or two possible candidates) for the infinitive, and none for the *-ing* participle (cf. Huddleston and Pullum et al. (2002: 112), who describe the stative idiom in these two forms as respectively very marginal and non-occurring).

For present purposes it is the overall pattern shown in Table 1 which is of interest. Our calculations show that this overall pattern is not altered by either (i) excluding all instances of the combination HAVE + *got*, or (ii) excluding estimated numbers of stative and ambiguous examples only. The two methods produce similar results (since few examples are clearly non-stative): they reduce the change in frequency for the present perfect category to around 0.8–0.9%, slightly reduce the change in the past perfect to about –34% (still a significant decline), and increase the change for the total to 10–11%.⁷

For the present perfect, our results for spoken English in DCPSE can be compared with those of Hundt and Smith (2009) for printed written English in the Brown quartet of corpora. They find a slight *decrease* in overall frequency, which is not statistically significant, for both British and American English, from the 1960s to the 1990s. Thus both studies find stable overall frequency levels. This does not, of course, rule out possible recent changes in the use and distribution of the present perfect.⁸ The existence of regional variation, often noted, is confirmed by Hundt and Smith's study, with the present perfect in American English starting from a lower level of frequency and remaining at a significantly lower level than in British English. Elsness (1997), who examined past-referring verb forms in his own corpora of historical and contemporary (1960s–1980s) material, also finds a significantly lower proportion of the present perfect in contemporary printed American than in British English, with the proportion in American English having fallen markedly between the period 1750–1800 and the contemporary period. In fact, Elsness, though focusing on the present perfect, also reports data showing regional variation in the past and infinitival perfect forms in contemporary printed English, with proportions of both forms again significantly lower in American than in British English, and again having fallen in American English since the 1750–1800 period.

Here our main concern is the contrasting trends in our frequency data for the present and non-present forms, with the stability of the former contrasting with the substantial declines in the past and infinitival forms. American influence may be a contributing factor in these declines, but this does not explain the contrasting patterns. The present perfect differs from the other forms in two ways which are likely to be relevant. First, it is much more frequent, and less frequent items may be more likely to suffer loss than more frequent ones, unless countervailing factors are involved (see e.g. Leech et al. 2009: 90, 269–70). Second, as noted above, the present perfect has a specialised pattern of use, involving an orientation towards present as well as past time, in contrast with the past tense. In any case, the DCPSE data show that the non-present perfect forms are worthy of investigation in terms of current change. The search capacities of ICECUP and the parsed corpus allow us to investigate further by examining their structural contexts of occurrence. The next section discusses the infinitival perfect (on the past perfect in DCPSE, see Bowie and Wallis, forthcoming).

⁷ See Bowie and Wallis (forthcoming) for further details of these results.

⁸ Genre may be an important dimension in this regard: for example, Hundt and Smith (2009) report some statistically significant changes in frequency over time for particular written genres, while Bowie and Wallis (forthcoming) note differences among spoken genres. There are also reports of particular uses of the present perfect as recent developments, such as use in the narration of past time events (e.g. Walker 2007 for British English).

3 The infinitival perfect in DCPSE

The infinitival perfect occurs in two main kinds of context: a bare infinitival construction with a preceding modal auxiliary (as in *we should **have brought** Dilys along*), or a *to*-infinitival construction (as in *she seems to **have been** far less tired*). A number of FTFs were constructed for these contexts. Figure 3 shows an FTF used to retrieve examples occurring within a VP after a modal auxiliary (the tree is displayed with branching from left to right, rather than from top to bottom). Note that a VP, in the parsing system used in the corpus, consists of the main verb and any preceding auxiliaries, with intervening material such as adverb phrases included (it does not include complements or adjuncts that follow the main verb). Intervening material is allowed for in the FTF by choosing the setting ‘next child: after’ rather than ‘next child: immediately after’, shown by the white arrow (so including examples like *might quite well **have died** in childbirth*). For the modal context, a second FTF (not shown) was used to find additional examples (far fewer in number) where the modal auxiliary preceded the subject and was therefore separated from the VP (in interrogatives such as *How old would you **have been***).

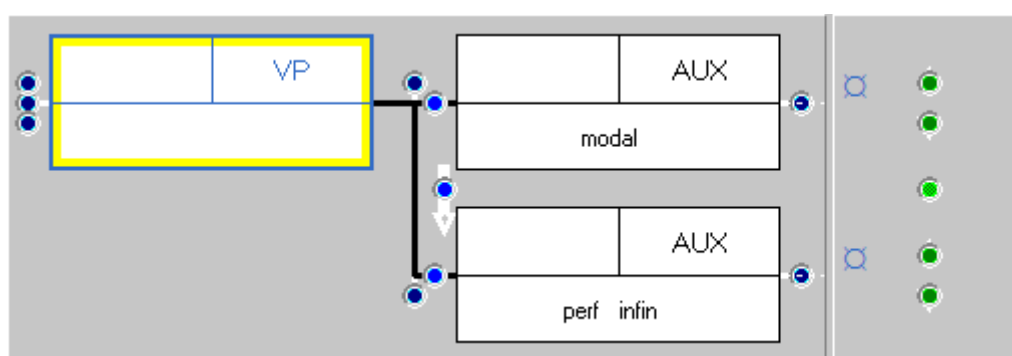


Figure 3: FTF for a perfect infinitive auxiliary following a modal auxiliary under a VP.

To-infinitival examples were found using several FTFs. The first pattern was as in Figure 3, but with the AUX feature ‘semi’ instead of ‘modal’. In the corpus the class of ‘semi-auxiliaries’ includes items such as BE *supposed to*, HAVE *to*, and SEEM *to* (discussed by Quirk et al. (1985: 141–7) as modal idioms, semi-auxiliaries, and catenatives). A second FTF looked for interrogative examples involving semi-auxiliaries, but no examples occurred in the corpus (a possible example would be *what was he supposed to **have done***). The third FTF, shown in Figure 4, searched in VPs following ‘particle’ *to* within a clause, and found examples within a variety of larger structures (e.g. *it’s nice to **have met** her*; *what would you claim to **have achieved***).

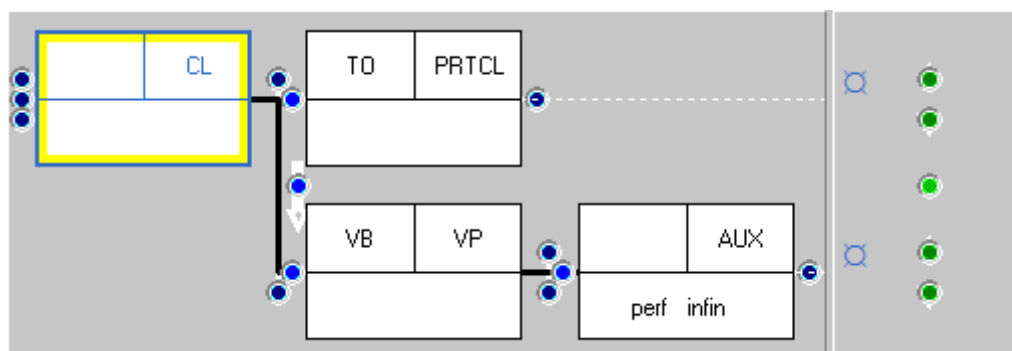


Figure 4: FTF for a perfect infinitive preceded by a *to*-particle under the same host clause.

The searches showed that, across the corpus, the great majority of examples of the perfect infinitive (88%) occur following a modal auxiliary. A decline in frequency has been observed

for the modals themselves in studies of recent change (e.g. Leech et al. 2009; Aarts, Bowie and Wallis forthcoming). In DCPSE, modal auxiliaries (which total 14,316) decline in frequency (as a proportion of words) by 6.4% from the earlier subcorpus to the later one (the result is significant at $p < 0.05$). This raises a question concerning the decline observed in the infinitival perfect across the two subcorpora of DCPSE: is it simply due to a decline in this major context of potential occurrence? To test this, use can be made of the same FTFs as described above, but with the node ‘AUX(perf, infin)’ omitted, in order to provide a total of potential contexts. The proportions in which a perfect infinitive occurs in any context, or combination of contexts, can then be calculated. The results for the modal and the *to*-infinitive contexts are shown in Tables 2a and 2b respectively.⁹

	perfect infinitive	no perfect infinitive	Total	Summary
LLC	561 (7.37%)	7,050	7,611	% change = -21.90 (c.i. \pm 11.20%)
ICE-GB	371 (5.76%)	6,074	6,445	
Total	932 (6.63%)	13,124	14,056	$2 \times 2 \chi^2 = \mathbf{14.69 \text{ s}}$

Table 2a: Changes in the proportion of perfect infinitives in modal contexts in the LLC and ICE-GB components of DCPSE. We find a ~20% fall from 7.37% to 5.76%; ‘c.i.’ = confidence interval.

	perfect infinitive	no perfect infinitive	Total	Summary
LLC	87 (1.33%)	6,447	6,534	% change = -51.83 (c.i. \pm 25.86%)
ICE-GB	40 (0.64%)	6,197	6,237	
Total	127 (0.99%)	12,644	12,771	$2 \times 2 \chi^2 = \mathbf{15.44 \text{ s}}$

Table 2b: Changes in the proportion of perfect infinitives in *to* contexts in the LLC and ICE-GB components of DCPSE. We find a ~50% fall from 1.33% to 0.64%; ‘c.i.’ = confidence interval.

These results show that the proportion of perfect infinitives has fallen significantly within both kinds of contexts. Therefore the overall decline in frequency of the infinitival perfect is not attributable solely to the decline in the frequency of modal auxiliaries, but involves independent trends of decline within possible contexts of occurrence. The decline within the *to* contexts is particularly steep, but involves much smaller numbers overall.

This does not in itself provide an explanation for the observed decline, as the presence and absence of a perfect infinitive in these contexts cannot in general be considered alternative choices for expressing very similar meanings. This is evident from pairs such as *he may be in London* versus *he may have been in London*, or *he is believed to be in London* versus *he is believed to have been in London*, where there is a clear temporal contrast. However, there are instances in the data where the use of a non-perfect variant would make no difference or only a subtle difference to the meaning conveyed. Consider the following examples.

(3) you know very well that your Party would have had to **have done** something uh if it had come back to power
[DL-D04 #41]

(4) there were to **have been** four greys in the field but the only one left is Marche d’Or three
[DL-F04 #54]

⁹ Note that the total of ‘modal contexts’ is slightly lower than the total of all modal auxiliaries; this is because the context FTFs exclude instances of modals where there is no associated VP node, as in tag questions or elliptical utterances like *could you*. A similar result is obtained if the set of all modal auxiliaries is chosen as the basis for comparison. Note also that there are 6 perfect infinitives from the total in Table 1 that are not accounted for by the FTF contextual searches.

(5) well I'd like you to **have found** out please

[DL-H01 #130]

(6) he's forty odd I would **have thought**

[DI-B49 #53]

In (3) the second perfect seems superfluous, as no further anteriority is involved; there are 12 such 'superfluous double perfect' examples in the corpus, of which 10 are from LLC. Such uses have been condemned by prescriptive grammarians from the 18th century onwards (Molencki 2003). In (4) (from a horse-racing commentary), where past tense *were* expresses past time, the perfect seems to convey non-actualisation of the situation, but as this is clear from the context it could have been left unexpressed (*there were to be four greys*, more plausible with stress on *were*). In (5), which conveys a directive, the perfect expresses a 'past in future' interpretation (the context suggests that the future reference point is 'before the next court hearing'), but again this could have been left unexpressed. In (6) the formulaic expression *I would have thought* is used as a 'hedge'; in most examples with such formulae, the perfect seems merely to add to the tentativeness expressed in *I would think*, perhaps because locating an opinion in past time suggests a readiness to revise it. Variants of the formula *{I/you/one} {would/should} have {thought/said}* are quite numerous in the corpus (70, allowing for negative and reduced forms of the auxiliaries; 47 of these are in the LLC subcorpus).¹⁰

There are also instances where there is a possible non-perfect variant with a preceding morphological past tense expressing past time, e.g.:

(7) and apart from that I mean my results are supposed to **have come** out today

[DI-C03 #237]

(8) well he's lucky to **have got** an extra hour in

[DL-A07 #41]

In (7) we might expect instead *my results were supposed to come out today* (cf. Collins 2009: 81–2), while (8) seems little different in meaning from *he was lucky to get an extra hour in*.

The examples discussed above suggest there is some leeway for the use of non-perfect variants instead of constructions with a perfect infinitive, so there may be an increasing tendency to simplify verb phrases where possible. However, this leeway seems to apply only in a restricted set of instances. It is hard to identify a determinate set of alternatives that would apply to other instances; this may require a broader-ranging investigation of the expression of modality in combination with past time reference. This is an area of English where there is considerable complexity in form–meaning mappings, which may lead to variation and instability as speakers reanalyse the mappings. Depraetere and Reed (2006: 287) comment that 'The area of temporal interpretation of modal utterances in English is one which is yet to be fully researched.' Research in this area from variationist and diachronic perspectives is also needed.

4 Conclusion

In this paper we have presented corpus evidence for contrasting trends in different subcategories of the perfect construction in spoken British English over recent decades: stable overall frequencies for the present perfect as against falling frequencies for the past and infinitival perfects. We have also examined the infinitival perfect in relation to its syntactic contexts of occurrence. The context following a modal auxiliary is by far the most frequent. Modals have themselves fallen in frequency; however, by taking into account the frequencies of the possible contexts of occurrence, evidence was provided of independent trends of decline in frequency of the perfect infinitive within these contexts.

¹⁰ It should be noted, however, that in a minority of these instances the perfect is not omissible. For example, one speaker, having expressed surprise to hear that it is raining, says *I would have thought it was too cold to rain*. Here it would not be appropriate to say *I would think it is/was too cold to rain*, as the evidence forces revision of the opinion.

This study has shown the importance of considering changes in a linguistic category like the perfect in relation to its interaction with other categories like tense and modality. The interaction of categories is likely to be important in change processes, especially in areas where form–meaning mappings are complex. The linguistic contexts of occurrence of a category can themselves change in frequency, and this needs to be taken into account—a process which is facilitated by use of a parsed corpus with a flexible facility for searching for structural patterns.

Acknowledgements

We are very grateful to Sean Wallis for assistance with the statistical analysis in this paper.

References

- Aarts, Bas, Jill Bowie and Sean Wallis (forthcoming) Profiling the English verb phrase over time: modal patterns. In: Merja Kyto and Irma Taavitsainen (eds.), xxx.
- Aarts, Bas, Joanne Close, Geoffrey Leech and Sean Wallis (eds.) *The English verb phrase: corpus methodology and current change*. Cambridge: Cambridge University Press.
- Aarts, Bas, Gerald Nelson and Sean Wallis (1998) Using Fuzzy Tree Fragments to explore English grammar. *English Today* 14: 3: 52–56.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999) *The Longman grammar of spoken and written English*. London: Longman.
- Bowie, Jill, and Sean Wallis (forthcoming) Changes in the perfect in spoken British English. In: Aarts, Close, Leech and Wallis (eds.).
- Collins, Peter (2009) *Modals and quasi-modals in English*. Amsterdam; New York: Rodopi.
- Davies, Mark (this volume) Why size matters: examining recent changes in medium and low-level syntactic phenomena with large corpora.
- Depraetere, Ilse, and Susan Reed (2006) Mood and modality in English. In: Bas Aarts and April McMahon (eds.), *The handbook of English linguistics* (pp. 269–290). Malden, MA; Oxford: Blackwell.
- Elsness, Johan (1997) *The perfect and the preterite in contemporary and earlier English* (Topics in English Linguistics, 21). Berlin; New York: Mouton de Gruyter.
- Huddleston, Rodney, and Geoffrey K. Pullum *et al.* (2002) *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hundt, Marianne, and Geoffrey Leech (this volume) Small is beautiful – on the value of standard reference corpora for observing recent grammatical change.
- Hundt, Marianne, and Nicholas Smith (2009) The present perfect in British and American English: has there been any change, recently? *ICAME Journal* 33, 45–63.
- Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith (2009) (eds.) *Change in contemporary English: a grammatical study*. Cambridge: Cambridge University Press.
- Molencki, Rafał (2003). Proscriptive prescriptivists: on the loss of the ‘pleonastic’ perfect infinitive in counterfactual constructions in Late Modern English. *Linguistic Insights: Studies in Language and Communication*, 7, 175–196.
- Nelson, Gerald, Sean Wallis and Bas Aarts (2002) *Exploring natural language: working with the British component of the International Corpus of English*. Amsterdam: Benjamins.
- Newcombe, R. G. (1998) Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 17: 873–890.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985) *A comprehensive grammar of the English language*. London: Longman.
- Sheskin, D. J. (1997) *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.
- Smith, Nicholas, and Geoffrey Leech (forthcoming) Verb structures in twentieth-century British English. In: Aarts, Close, Leech and Wallis (eds.).

- Svartvik, Jan (ed.) (1990) *The London-Lund Corpus of spoken English: description and research*. Lund Studies in English 82. Lund: Lund University Press.
- Walker, Jim (2007) The strange phenomenon of the footballer's perfect. MS Université Lumière Lyon 2.