

Change in the English infinitival perfect construction

Jill Bowie and Bas Aarts

[pre-publication version, February 2011]

1 Introduction

The availability of searchable electronic corpora composed of textual material from different time periods has made studying change in the English language easier. However, 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 2012). While acknowledging the distinct advantages of using large corpora, we believe that detailed analysis of small corpora, especially those which are parsed, can reveal trends that may be missed by other approaches. Furthermore, spoken language corpora are particularly valuable for studying short-term change: spoken language is primary, and grammatical changes are likely to manifest themselves in that medium first.¹

The study of recent grammatical change presented here draws on the Diachronic Corpus of Present-Day Spoken English (DCPSE), based at the Survey of English Usage, UCLDCPSE is unique in several ways. First, it contains exclusively spoken (and mainly spontaneous) English, in two subcorpora with matching text categories, allowing diachronic comparison over a thirty-year span. The earlier subcorpus contains approximately 464,000 words from the London–Lund Corpus (LLC) from the late 1950s–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).²

Our study focuses on the British English (BrE) perfect construction, and in particular the infinitival perfect. We first compare the inflectional subtypes of the perfect (present/past/non-finite) in terms of changing frequencies of use, before examining the infinitival perfect in more detail. There are several reasons for investigating the perfect. The first is to extend what is known about its longer-term historical development. From early origins in Old English, the perfect increased markedly in frequency through Middle English into early Modern English, but this advance later appears to have been halted and even, after 1800, reversed, at least in American English (AmE) (Elsness 1997; Fischer and van der Wurff 2006). Second, contemporary written

¹ We are very grateful to Sean Wallis for assistance with the statistical analysis in this paper. We also acknowledge with gratitude the support of the Arts and Humanities Research Council under grant AH/E006299/1.

² See Svartvik (1990) on LLC and Nelson, Wallis, and Aarts (2002) on ICE-GB/ICECUP; see also <http://www.ucl.ac.uk/english-usage>

sources show that the perfect is less frequent in AmE than in BrE (Elsness 1997; Hundt and Smith 2009);³ it is therefore worth investigating whether British usage is changing towards American norms as appears to have happened with other grammatical features (Leech et al. 2009), since the spoken language may be “ahead” in this regard. Finally, the perfect provides a case study of a construction whose subtypes show distinctive patterns in terms of relative frequency, semantic specialisation, and syntactic contexts of occurrence. Such factors are likely to influence changing patterns of usage, making a fine-grained analysis valuable.

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 anteriority to a past reference point, and the two non-finite subtypes (infinitival and *-ing*-participial) indicate 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 other perfect subtypes, which can correspond to either a present perfect or a simple past (Huddleston and Pullum 2002: 146).

Examples of the perfect construction can be retrieved in DCPSE by using Fuzzy Tree Fragments (FTFs), a search facility within ICECUP which allows users to construct partial tree diagrams and to choose the level of detail specified (Aarts, Nelson, and Wallis 1998; Nelson, Wallis, and Aarts 2002). Figure 1 shows a simple FTF search for a single node of category “auxiliary” with type feature “perfect” and inflectional feature “infinitive”. Categorical information (e.g. word or phrase class) is displayed in the upper righthand segment, functional information (such as subject, NP head) in the top lefthand segment, and additional features in the lower segment. In this instance the function has been left unspecified.

³ Hundt and Smith (2009) look at the present perfect; Elsness (1997) also looks at other perfect subtypes.

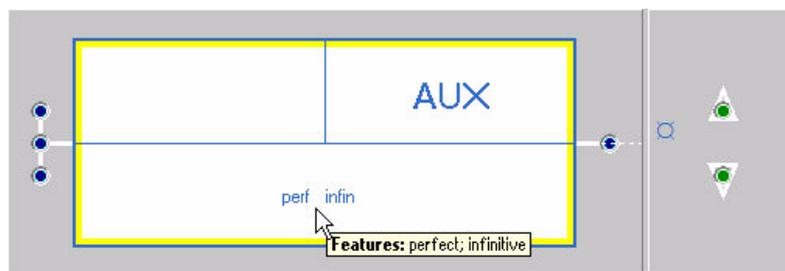


Figure 1. FTF for perfect infinitive auxiliaries

The combination HAVE+*got* requires special attention. “Semi-modal” HAVE *got* [*to*] (as in *a lot of work has got to be done*) is unproblematic: it is analysed in the corpus as an auxiliary with type feature “semi” and so automatically excluded by FTF searches for perfect auxiliaries. However, the searches do include, alongside clear instances of the perfect construction with *got* (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’). These represent an idiom historically derived from a perfect construction but now semantically distinct (Huddleston and Pullum 2002: 111–13). Although they are sometimes included in corpus counts of the perfect (e.g. Biber et al. 1999: 463–67), they are best excluded. 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, comprising around 24% of examples – examination of a 10% random sample showed that a majority are stative or ambiguous, so all instances were excluded from the counts.⁴ Occurrences with *got* were far less frequent in the other inflectional categories; all examples were examined, and stative and ambiguous ones excluded from the counts (necessary only for the past perfect).

Frequencies of the perfect (normalised per million words, “pmw”) were then compared for LLC and ICE-GB. The results show that the perfect auxiliary falls in frequency by 10.59 % across the two subcorpora (Table 1, Total row).⁵

⁴ This does exclude some genuine perfect examples. An alternative calculation excluded only stative and ambiguous examples involving *got* for the present tense category, based on estimations from the random samples. This produced very similar results to those in Table 1, since few examples are clearly non-stative.

⁵ Note that the data reported in this paper are based on a revised version of DCPSE prepared at the Survey of English Usage.

Table 1. Frequencies of perfect auxiliaries in DCPSE, divided by inflectional category. Columns A and B represent goodness of fit ² comparisons summarised in the text. Results marked ‘s’ are significant at p<0.05.

Note: Figures exclude certain instances of HAVE *got* (see discussion in text).

Inflectional category	LLC		ICE-GB		Change in frequency		
	raw	pmw	raw	pmw	%	A: χ^2 (words)	B: χ^2 (perfect)
present	2,759	6,194.75	2,523	6,241.22	0.75%	0.07	18.82 s
past	787	1,767.04	473	1,170.07	-33.78%	50.92 s	26.84 s
infinitive	652	1,463.93	416	1,029.07	-29.70%	31.88 s	14.76 s
-ing participle	78	175.13	58	143.48	-18.08%	1.33	0.25
Total	4,276	9,600.86	3,470	8,583.84	-10.59%	24.04 s	

However, Table 1 also shows that not every category behaves in the same manner. Past and infinitival forms fall by around 34% and 30% respectively, whereas the present (by far the most frequent of the forms) is stable and indeed slightly increases in real terms. We report 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.⁶ Note that the slight percentage increase (0.75%) of the present tense category 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 95% confidence intervals depicted by error bars.⁷

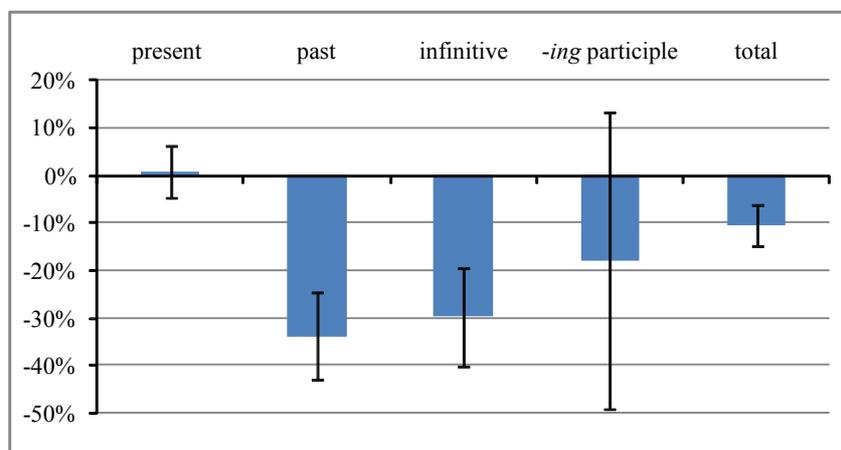


Figure 2. Changes in pmw frequencies (Table 1 “%” column) with error bars for p<0.05. Where an error bar does not cross the zero axis, the change is statistically significant (cf. Table 1 Column A).⁸

⁶ To be more precise, in Column A we carry out a goodness-of-fit ² test (Sheskin 1997: 95) for the overall change (in the Total row) and for each individual subcategory (present, past, etc.) 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.

⁷ 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.

⁸ This type of visualisation displays the size of the result (i.e. the column height) and our confidence in it. For example, we are 95% certain that the “past perfect” category falls by between 25% and 43%.

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 pmw frequency, which is not statistically significant, for both BrE and AmE, from the 1960s to the 1990s (with the frequency in AmE starting lower and remaining significantly lower than in BrE). They also consider the relative proportions of present perfect and morphological past tense, and again find a pattern of overall stability. Further analysis of our data gives a contrasting result: the present perfect actually shows a significant proportional increase (around 12%) against the morphological past tense.⁹ Thus the present perfect deserves further investigation in terms of possible changes in use and distribution.¹⁰ Genre variation is likely to be relevant here (cf. Gries 2006); Hundt and Smith (2009) report some differing diachronic trends in pmw frequency for particular written genres. As the two parts of DCPSE consist of material in a range of matched spoken genres, it is possible to explore further by comparing diachronic trends in these genres.

Here, however, our main concern is the substantial declines observed for the past and infinitival subtypes, in marked contrast with the findings for the present perfect. American influence may be a contributing factor in these declines. Elsness (1997) reports data on the past perfect and infinitival perfect in printed English, showing that proportions of both (within the set of past-referring forms) are significantly lower in American than in British contemporary material, having fallen in AmE since 1800.¹¹ However, he finds the same pattern for the present perfect, which is not declining in our data. Two factors may help to explain the contrast. First, as noted above, the present perfect differs from the other perfect subtypes in having a specialised pattern of use, involving an orientation towards present as well as past time. Second, the other subtypes are much less frequent than the present perfect, and less frequent items may be more likely to suffer loss than more frequent ones (see e.g. Leech et al. 2009: 90, 269–70). The next section discusses the infinitival perfect in more detail (on the past perfect in DCPSE, see Bowie, Wallis, and Aarts, forthcoming 2012).

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 (*we should **have brought** Dilys along*), or a *to*-infinitival construction (*she seems to **have been** far less tired*). Several FTFs were constructed for these contexts. Figure

⁹ For further details, see Bowie, Wallis, and Aarts (forthcoming 2012), which also reports a significant proportional decline against the morphological past tense for both the past perfect and the category modal+perfect infinitive.

¹⁰ An interesting narrative use of the present perfect has been reported for recent BrE, although it may represent a longer-standing non-standard use (Walker 2008).

¹¹ This is based on our calculations combining figures provided by Elsness for several different constructions involving these perfect subtypes (1997: 104, 267–68).

3 shows an FTF (with left-to-right branching) to retrieve examples occurring within a VP after a modal auxiliary. In the corpus, a VP consists of the main verb and any preceding auxiliaries, with intervening material such as adverb phrases included. 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*, and even *that must I fancy from the way he played it have come back onto him very very sharply*).

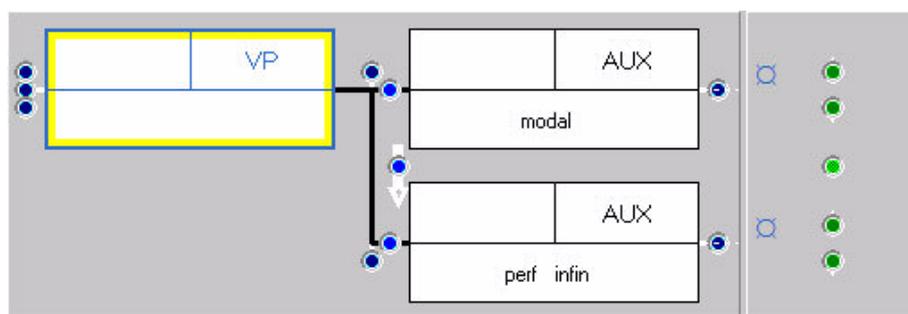


Figure 3. FTF for a perfect infinitive auxiliary following a modal auxiliary under a VP

For the modal context, a second FTF retrieved additional examples where the modal auxiliary preceded the subject and was therefore separated from the VP (in interrogatives such as *How old would you have been?*). Several further FTFs were used to find *to*-infinitival examples occurring in structures parsed as involving “semi-auxiliaries” (e.g. BE *supposed to*, HAVE *to*, SEEM *to*) and those parsed as clauses introduced by “particle” *to* (e.g. *What would you claim to have achieved specifically?*).

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 2012). 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: is it simply due to a decline in this major context of potential occurrence? This was tested by (i) using the same FTFs as described above, but with the node “AUX(perf, infin)” omitted, to quantify potential contexts; and (ii) calculating the proportions of such contexts in which a perfect infinitive occurs. The results for modal and *to*-infinitive contexts are shown in Tables 2a and 2b respectively.¹²

¹² The total of “modal contexts” is slightly lower than the total of all modal auxiliaries, because the context FTFs exclude instances of modals with no associated VP node, as in tag questions or elliptical utterances like *could you*. A

Table 2a. Changes in the proportion of perfect infinitives in modal contexts in the LLC and ICE-GB components of DCPSE. The result is significant ('s') for $p < 0.05$.

	perfect infinitive	no perfect infinitive	Total	² score
LLC	561 (7.37%)	7,050	7,611	% change = 21.39 (c.i. \pm 11.10)
ICE-GB	374 (5.79%)	6,081	6,455	
Total	935 (6.65%)	13,131	14,066	14.00 s

Table 2b. Changes in the proportion of perfect infinitives in *to* contexts in the LLC and ICE-GB components of DCPSE. The result is significant ('s') for $p < 0.05$.

	perfect infinitive	no perfect infinitive	Total	² score
LLC	87 (1.33%)	6,447	6,534	% change = -51.86 (c.i. \pm 25.63)
ICE-GB	40 (0.64%)	6,201	6,241	
Total	127 (0.99%)	12,648	12,775	15.47 s

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 explain 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 a non-perfect variant would differ in meaning only subtly, at most:

- (1) 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]¹³
- (2) there were to **have been** four greys in the field but the only one left is Marche d'Or three [DL-F04 #54]
- (3) well I'd like you to **have found** out please [DL-H01 #130]
- (4) he's forty odd I would **have thought** [DI-B49 #53]

similar result is obtained if the set of all modals 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.

¹³ The prefixes "DL" and "DI" indicate examples from the LLC and ICE-GB subcorpora of DCPSE.

In (1) the second perfect seems superfluous, as no further anteriority is involved; of 12 such “superfluous double perfect” examples in the corpus, 10 are from LLC. Such uses have been condemned by prescriptive grammarians from the eighteenth century onwards (Molencki 2003). In (2) (from a horse-racing commentary), where past tense *were* expresses anteriority, 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 (3), which conveys a directive, the perfect expresses a “past-in-future” interpretation (the context suggests that the future reference point is the next court hearing), but again this could have been left unexpressed. In (4) 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.¹⁴ 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).¹⁵

In other instances there is a possible non-perfect variant with a preceding morphological past tense expressing past time:

(5) and apart from that I mean my results are supposed to **have come** out today [DI-C03 #237]

(6) well he’s lucky to **have got** an extra hour in [DL-A07 #41]

In (5) we might expect instead *my results were supposed to come out today* (cf. Collins 2009: 81–82), while (6) 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; beyond these, non-perfect alternatives would involve quite different structures (such as *perhaps I missed something* instead of *I may have missed something*), making it hard to identify a determinate set of alternatives. This may require a broader-ranging investigation of the expression of modality in combination with past time reference. This area of English involves considerable complexity in form–meaning mappings, which may lead to variation and instability as speakers reanalyse the mappings. Depraetere and Reed (2006: 287) note the need for more

¹⁴ However, in a minority of 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 *I would think it is/was too cold to rain* would not be appropriate, as the evidence forces revision of the opinion.

¹⁵ These perfect and non-perfect formulae can also be considered as a separate set (n=222). Variants of the non-perfect formulae *I would think/say* are also more frequent in LLC than in ICE-GB; the proportion of formulae which are perfect is slightly higher in LLC (32% vs. 30%), but the difference is not significant. Removing the set of formulae from Table 2a makes little difference to the results.

research on the temporal interpretation of modal utterances in English; this should include research from variationist and diachronic perspectives.

4 Conclusion

Our corpus data for spoken BrE reveals contrasting trends in different subcategories of the perfect construction over recent decades: the past and infinitival perfects decline, while the present perfect is stable in pmw frequency and indeed increases proportionally against the past non-perfect. Factors behind the declines may include American influence and a tendency to simplify the VP. The different pattern for the present perfect may be due to its much greater frequency and to its semantic specialisation (involving reference to present as well as past time).

Closer examination of the infinitival perfect shows that it most frequently occurs following a modal auxiliary. Modals have themselves fallen in frequency; however, by taking into account the frequencies of possible contexts of occurrence (i.e. modal and *to*-infinitival structural contexts), independent trends of decline of the perfect infinitive within these contexts can be shown.

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 morphological tense and modality. The interaction of categories is likely to be important in change processes, especially in areas where form–meaning mappings are complex. Our study has also highlighted the need to consider the relative frequencies of these interacting categories. The investigation of such complexities is facilitated by use of a parsed corpus like DCPSE with a flexible means of searching for structural patterns.

References

- Aarts, Bas, Jill Bowie, and Sean Wallis. Forthcoming 2012. 'Profiling the English Verb Phrase over Time: Modal Patterns'. In *Corpus Linguistics and the History of English*, ed. Merja Kytö and Irma Taavitsainen. Cambridge: Cambridge University Press.
- Aarts, Bas, Gerald Nelson, and Sean Wallis. 1998. 'Using Fuzzy Tree Fragments to Explore English Grammar'. *English Today* 14: 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, Sean Wallis, and Bas Aarts. Forthcoming 2012. 'Changes in the Perfect in Spoken British English'. In *The English Verb Phrase: Corpus Methodology and Current Change*,

- ed. Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Cambridge: Cambridge University Press.
- Collins, Peter. 2009. *Modals and Quasi-Modals in English*. Amsterdam: Rodopi.
- Depraetere, Ilse and Susan Reed. 2006. 'Mood and Modality in English'. In *The Handbook of English Linguistics* ed. Bas Aarts and April McMahon, 269–90. Malden, MA: Blackwell.
- Elsness, Johan. 1997. *The Perfect and the Preterite in Contemporary and Earlier English. Topics in English Linguistics, 21*. Berlin: Mouton de Gruyter.
- Fischer, Olga and Wim van der Wurff. 2006. 'Syntax'. In *A History of the English Language*, ed. Richard Hogg and David Denison, 109–98. Cambridge: Cambridge University Press.
- Gries, Stefan Th. 2006. 'Exploring Variability Within and Between Corpora: Some Methodological Considerations'. *Corpora* 1: 109–51.
- Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- 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, eds. 2009. *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: 175–96.
- 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, Robert G. 1998. 'Interval Estimation for the Difference between Independent Proportions: Comparison of Eleven Methods'. *Statistics in Medicine* 17: 873–90.
- Sheskin, David J. 1997. *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton: CRC Press.
- Smith, Nicholas and Geoffrey Leech. Forthcoming 2012. 'Verb Structures in Twentieth-Century British English'. In *The English Verb Phrase: Corpus Methodology and Current Change*, ed. Bas Aarts, Joanne Close, Geoffrey Leech, and Sean Wallis. Cambridge: Cambridge University Press.
- Svartvik, Jan, ed. 1990. *The London-Lund Corpus of Spoken English: Description and Research*. Lund: Lund University Press.
- Walker, Jim. 2008. 'The Footballer's Perfect: Are Footballers Leading the Way?' In *The Linguistics of Football*, ed. Eva Lavric, Gerhard Pisek, Andrew Skinner, and Wolfgang Stadler, 295–303. Tübingen: Narr.

Corpora used:

Diachronic Corpus of Present-Day Spoken English (DCPSE), Survey of English Usage, UCL
<http://www.ucl.ac.uk/english-usage/projects/dcpse/>.