The perfect in spoken British English[1]

Jill Bowie, Sean Wallis and Bas Aarts
University College London

1   Introduction

The English perfect construction involves the perfect auxiliary HAVE followed by a verb in the past participle form. It occurs in several subtypes according to the inflectional form of the auxiliary. The most frequently occurring is the present perfect, as in *She **has seen** them*. The other subtypes are the past perfect (*She **had seen** them*), the infinitival perfect (*She must **have seen** them*), and the *-ing*-participial perfect (***Having seen** them, she went home*). All subtypes typically function to express anteriority (i.e. pastness relative to a reference point), although further semantic complexities have led to varying treatments of the perfect, for example as an aspect or a secondary tense system.

Research on the English perfect has revealed considerable variation in use both diachronically, across longer historical periods, and synchronically, across regions and dialects. Recent trends in perfect usage are therefore of interest. The study presented here investigates this topic with regard to spoken standard British English. Much previous work has focused mainly or exclusively on the present perfect. However, here we investigate all inflectional subtypes of the perfect in terms of frequency changes over time. The findings on the present perfect are compared with those of other researchers. We then focus in more detail on the past perfect and infinitival perfect, to seek explanations for the frequency changes observed.

In the remainder of this introduction we elaborate on the reasons for undertaking this study (1.1) and introduce the corpus used in our research (1.2).

1.1   Reasons for this study

There are several reasons for investigating recent change in the perfect in spoken British English. These concern the longer-term history of the perfect in English, regional variation in the perfect, and the need for research on the spoken language.

First, the longer-term history of the perfect makes it of interest to investigate current trends. From early origins in Old English, it increased markedly in frequency through Middle English into early Modern English, coming into competition with the morphologically marked past tense; but there is some evidence that this advance has more recently been halted or even reversed, especially in American English (Elsness 1997; Fischer and van der Wurff 2006). A corpus-based study of the present perfect by Elsness (1997), sampling the language at 200-year intervals, reports frequencies which increase sharply from Old English to 1600, before levelling off to 1800, and thereafter showing a clear decrease in American English to the present day. The trend from 1800 to the present day is less clear for British English, with some variation according to text category, but Elsness suggests the overall pattern is also one of declining frequency. If this is correct, it contrasts with the pattern found for a number of other European languages, where the grammaticalisation process has continued, leading the present perfect to encroach further on the territory of the morphological past tense or even replace it entirely, as in southern German (Heine and Kuteva 2006: 36–42).

However, there are contrasting suggestions in the literature that the present perfect is expanding its territory in British English, at least in some varieties: there have been reports of narrative uses (e.g. Walker 2008), and observations of apparently increasing use with adverbials (adjuncts) indicating specific past time reference (e.g. Hughes *et al*. 2005: 12–13).[2] It is therefore of interest to further investigate current developments in the use of the perfect.

[2] Use of the present perfect with past time specification is not in fact new: it occurred, apparently more freely, in earlier periods of (British) Modern English (e.g. Elsness 1997: 292–3, 2009a: 235–6). It appears to be relatively frequent in

Second, there is evidence of synchronic regional variation, with a number of studies showing a lower frequency of the present perfect in American than in British English (e.g. Elsness 1997; Hundt and Smith 2009).[3] It is therefore worth investigating whether British English usage is changing under the influence of American English in this area as appears to be the case for some other areas of grammar (e.g. Leech *et al*. 2009). Elsness (1997), 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 1800.[4] Gorrell (1995) cites examples from recent American English which suggest there may be further decline in the past perfect in this variety. These points indicate that the past and infinitival forms are also worthy of investigation in terms of current change.

Third, there is a need for more research on spoken English. In exploring short-term change, it is particularly valuable to look at spoken language, where changes in grammar are likely to first become evident. For the written language, recent change in the present perfect has been investigated by Hundt and Smith (2009), based on the 'Brown quartet'. These are four one-million-word corpora of printed English: for British English, *LOB* and *FLOB* (containing material from 1961 and 1991 respectively), and for American English, *Brown* and *Frown* (1961 and 1992). The present study draws on data from a corpus of spoken British English which covers a similar time period, introduced in the next section.

In this study we aimed to answer the following research questions:

- Does the frequency of use of the perfect change significantly in spoken standard British English from the 1960s to the 1990s?
- How do the subtypes of the perfect compare in this regard?
- What are the explanatory factors behind any frequency changes found?

We also decided to compare different baselines for measuring change over time, going beyond per-million-word measures to what we will argue are more informative measures, calculated as a proportion of VPs or past-marked VPs.

1.2   The corpus

The data in this study is drawn from the *Diachronic Corpus of Present-day Spoken English* (DCPSE), a parsed corpus of mainly spontaneous British English speech (described in more detail in Aarts *et al*., this volume). It comprises two subcorpora of over 400,000 words each in matching text categories, allowing diachronic comparison over a thirty-year span. The subcorpora contain material from (i) the *London–Lund Corpus* (LLC) dating from the late 1950s to the 1970s, and (ii) the *British Component of the International Corpus of English* (ICE-GB) collected in the early 1990s.

DCPSE is fully parsed in the form of phrase structure tree diagrams, and is searchable with dedicated corpus exploration software called ICECUP 3.1 (*International Corpus of English Corpus Utility Program 3.1*).[5] An example of a tree diagram for a sentence from the corpus is shown in Figure 1.

---

contemporary spoken Australian English (Engel and Ritz 2000; Elsness 2009b), for which narrative uses have also been reported (Engel and Ritz 2000).

[3] There has also been considerable research on the perfect in a number of other contemporary and earlier varieties of English (e.g. Tagliamonte 2000 on Samaná English, spoken by a small community in the Dominican Republic; Siemund 2004 on Irish English; van Herk 2008 on early African American English). A review of this literature is beyond the scope of this paper.

[4] For the infinitival perfect, we performed our own calculations based on combinations of the figures provided by Elsness for several different constructions involving this form (pp. 104, 267–8).

[5] See Svartvik (1990) on LLC and Nelson *et al.* (2002) on ICE-GB and ICECUP; for more information on DCPSE, see www.ucl.ac.uk/english-usage/projects/dcpse.
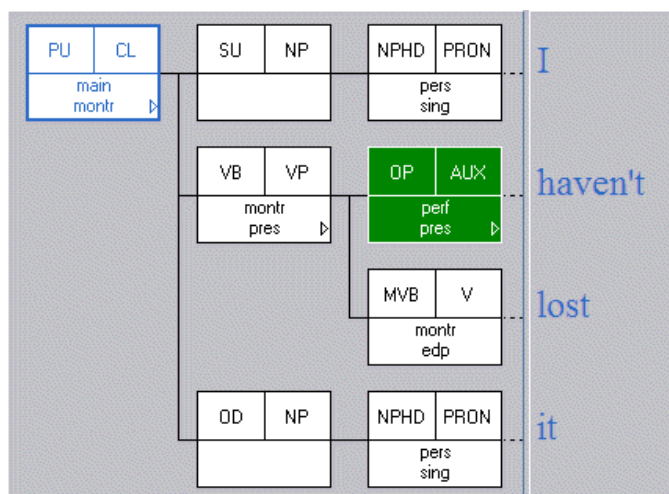

Figure 1. Tree diagram for the sentence *I haven't lost it*.[6]

The tree is displayed here branching from left to right. Each node of the tree has three sections: the upper right section shows categorial information (such as 'NP' for noun phrase, 'PRON' for pronoun), the upper left section displays functional information (such as 'SU' for subject, 'NPHD' for noun phrase head), and the lower section shows additional features (such as 'montr' for monotransitive, 'pres' for present tense).

ICECUP provides the facility to search for grammatical structures by constructing *Fuzzy Tree Fragments* or FTFs (Aarts *et al*. 1998; Nelson *et al*. 2002).[7] FTFs are a kind of 'wild card' for grammar: partial tree diagrams in which varying levels of detail can be specified that will then match the same configuration in trees. Figure 2 shows a single-node FTF designed to find every instance of a present-tense perfect auxiliary. Figure 3 shows a more complex FTF, in the form of a mini tree, which searches for every VP containing an infinitival perfect auxiliary preceded by a modal auxiliary verb.
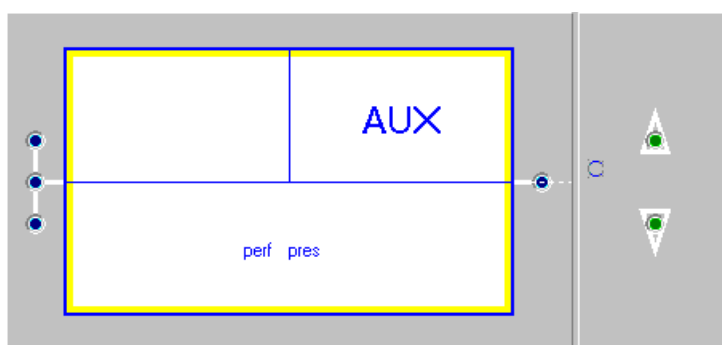

Figure 2. A simple FTF to search for a present perfect auxiliary.

Note that, in the phrase structure grammar used in DCPSE, a VP consists of the lexical verb together with any accompanying auxiliaries: that is, of a verbal group excluding such elements as objects and adverbials which occur after the lexical verb (but including any interposed adverbials, as in *It might well have been*).[8] Where an auxiliary is separated from the rest of the VP (e.g. in an interrogative, where the first auxiliary precedes the subject), it is treated as a daughter of the clause.

---

[6] Gloss (features are in italics): PU = parsing unit; CL = clause; *main* = main; *montr* = monotransitive; SU = subject; NP = noun phrase; NPHD = noun phrase head; PRON = pronoun; *pers* = personal; *sing* = singular; VB = verbal; VP = verb phrase; *pres* = present (tense); OP = operator; AUX = auxiliary; *perf* = perfect; MVB = main verb; V = verb; *edp* = -*ed* participle; OD = direct object.

[7] See also www.ucl.ac.uk/english-usage/resources/ftfs.

[8] There are of course other analyses of the VP in the linguistic literature; here we are not concerned with competing analyses, but with practical issues of data retrieval.
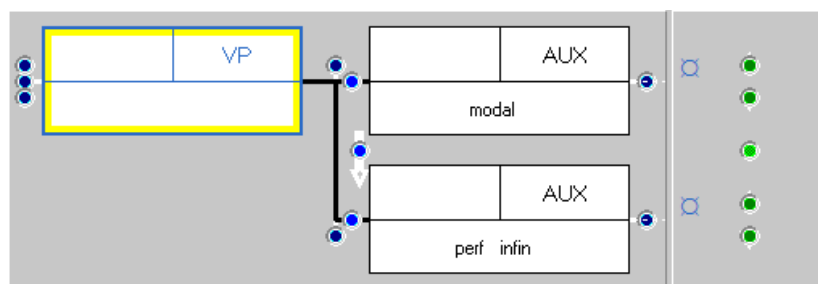

Figure 3. FTF for an infinitival perfect auxiliary following a modal auxiliary under a VP. The white arrow requires that the second auxiliary follows the first, but not necessarily in strict succession.

While a parsed corpus like DCPSE offers many advantages in studying grammatical structures, it is always important to check the results of electronic searches for accuracy and completeness. The design of ICECUP facilitates an exploratory mode of working with the data so that the user can readily confirm results by inspecting matching sentences.

In the next section of the paper we present data for frequency changes in the perfect construction across the two subcorpora of DCPSE, comparing trends for the construction as a whole and for the different inflectional subtypes. The trends seen for the past perfect and infinitival perfect are then taken up in more detail in section 3. Finally, we present our conclusions in section 4.

2   Subtypes of the perfect: frequency trends

As mentioned above, the perfect construction occurs in present, past and non-finite forms, all of which typically express anteriority to a reference point. Before presenting our quantitative findings, we survey these subtypes of the perfect in a little more detail, providing examples from DCPSE. Our description draws on accounts in the literature, including standard grammars such as Quirk *et al*. (1985) and Huddleston and Pullum *et al*. (2002).

Examples of the present perfect and past perfect are given in (1) and (2) respectively:[9]

(1)   *I've done a bit of writing before* [DI-A07 #176]
(2)   *I was staying in this sort of bedsitter place that she'd booked up for me*
[DL-B19 #655]

The present perfect (with auxiliary HAVE in the present tense) typically expresses anteriority relative to the present: in (1) we understand that the writing was done 'before now'. The past perfect (with the auxiliary in the past tense) typically indicates anteriority to a past reference point: in (2), we understand that the 'staying' was in the past, and the 'booking up' happened before the staying.[10]

The two non-finite subtypes of the perfect are illustrated below. The first is the infinitival perfect: the auxiliary either occurs as a bare infinitive following a modal auxiliary, as in (3), or follows the infinitival marker *to*, as in (4). The second non-finite subtype is the *-ing*-participial perfect, as in (5).

(3)   *I must have told you that* [DL-B19 #271]
(4)   *the virus seems to have cleared up* [DI-A10 #140]
(5)   *Her daughter Carol having produced coffee for the waiting reporters offered some thoughts on the matter* [DI-J04 #108]

The non-finite subtypes also indicate anteriority, with the reference point determined by the larger linguistic context in which they occur (Huddleston and Pullum *et al*. 2002: 140, 161–2). For example,

---

[9] Examples from DCPSE are cited with their identifying text codes (prefixes 'DL' and 'DI' indicate LLC and ICE-GB subcorpora respectively).
[10] Other uses of the past perfect (in backshifting and modal remoteness contexts) are discussed in section 3.

in (4) the 'clearing up' is understood as happening before the present time referred to by *seems*, while in (5) the 'producing coffee' is understood as happening before the past time at which the 'offering of thoughts' occurred.

The meaning of the present perfect is, however, more complex than anteriority to a present reference point. It contrasts with the (morphologically marked) past tense in ways which are difficult to characterise (for a detailed discussion, see Elsness 1997). Briefly, the present perfect presents a situation as occurring within (or even continuing through) a timespan beginning in the past and leading up to the present. It also typically involves a focus on the present repercussions of the situation (often labelled 'current relevance'), and generally resists co-occurrence with expressions indicating a specific past time reference (such as *last year*). In contrast, the past tense is readily used to refer to a time viewed as wholly in the past rather than connected to the present. For example, *he's lost his job* tends to highlight the current state of affairs (his joblessness) resulting from the past event, while *he lost his job* is compatible with reference to a particular time in the past (such as last year) and could occur as part of a narrated sequence of past events. Where the situation continues up to the present, the present perfect is generally required (*They've known him since 1983* vs *They knew him since 1983*). Admittedly, the distribution of the two forms is not altogether clear-cut, and varies somewhat across varieties of English, so these distinctions are sometimes blurred.[11]

The contrast between present perfect and past tense is neutralised in the other perfect subtypes (i.e. the past perfect and non-finite 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 (6) corresponds to a present perfect (with *since the election* specifying a period up to the speaker's present) and (7) to a simple past (as a specific time in the past is under discussion). That is, the content of the claim in (6) would be rendered as 'We have achieved *x* since the election', that in (7) as 'I heard a great deal of noise ...').

> (6)   *tick off for me the main things that you would claim to **have achieved** as a government since the election* [DL-E02 #27]
> (7)   *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 #99]

Having identified subtypes of the perfect, we can now attempt to explore how their usage may change over time. Aarts *et al.* (this volume) discuss three different baselines for measuring frequencies and evaluating change over time. These are: *simple word frequency*, frequency of a *higher-order grammatical term* (e.g. VP frequency for progressive verbs), and the frequency of members of a *set of alternants* containing the item. Aarts *et al.* argue that, wherever possible, one should employ a baseline for comparison that is closest to the set of *choices* that speakers are making between forms as this will provide the most informative results.

In the remainder of this section we compare the results of several different methods as applied to subtypes of the perfect. First, frequencies of perfect auxiliaries per million words are reported. Next, the perfect is considered as a choice (a) within the set of VPs, and (b) within the set of tensed, past-marked VPs (defined as tensed VPs marked for past either inflectionally or with the perfect, or both).

---

[11] For example, the following kinds of uses have been reported in the literature: past tense with current relevance, especially in American English (*He (just) **lost** his job*; ***Did** you ever **visit** Rome?*); past tense with *always* for a state continuing up to the present (*I want to see what happens next—I always **liked** a good story for a starter, and I like to see how they finish off* [DI-D06#219–20]; cf. Elsness 1997: 132; Leech 2004: 43); present perfect with past time adverbial (*oh **I've got** [='navigated'] them through there ages ago* [DL-B11#737]). Several studies note variation in the use of the perfect among different regional and/or non-standard varieties of British English (e.g. Sampson 2002; Miller 2004).

2.1   Perfect auxiliaries per million words

2.1.1   Data retrieval

Instances of the perfect auxiliary were retrieved by using a simple FTF search for a single node of category 'AUX' with the auxiliary type feature 'perf'. Different values of the inflectional 'tense/mood/form' feature were added in turn to this node to carry out more specific searches for perfect auxiliaries occurring in present tense, past tense, or one of the non-finite forms. An FTF used to search for perfect auxiliaries in the present tense form is shown in Figure 2 above.

The frequent combination HAVE + *got* required special attention: while some instances clearly belong to the perfect construction (e.g. *How advanced have they got* [DI-B24 #63]), the majority do not. Instead they represent an idiom which is historically derived from a perfect construction but is now semantically distinct, expressing various stative meanings (Huddleston and Pullum *et al.* 2002: 111–13). Occurrences of this idiom as 'semi-modal' HAVE *got [to]* (as in *a lot of work has got to be done on it* [DL-A04 #265]) were unproblematic: these were automatically excluded by our FTF searches, as it is analysed in the corpus as a semi-auxiliary (an auxiliary with the type feature 'semi'). However, our searches did include instances of the idiom taking an NP object and expressing a stative meaning such as possession (or a similar relationship): an example is *Well he's got two kids for one thing* [DI-B63 #124] (normally used to describe a present state of affairs, 'he has two kids', rather than to mean 'He has obtained two kids').[12]

We therefore used FTFs to find instances where HAVE, parsed as a perfect auxiliary, was followed by *got* (allowing for intervening material such as adverbs). In our spoken data, these were very frequent in the present tense category, contributing around 24% of examples; examination of 10% random samples of occurrences (from each of the subcorpora, LLC and ICE-GB) showed that a majority of examples were stative or ambiguous, so all instances were excluded from the counts for the present subtype of perfect.[13] Occurrences with *got* were far less frequent in the other inflectional categories (i.e. past, infinitive, and *-ing* participle); all examples were examined, and stative and ambiguous ones excluded from the counts (necessary only for the past category).

2.1.2   Results

Table 1 analyses the results using per million words ('pmw') as a baseline, contrasting LLC (the earlier subcorpus) and ICE-GB (the later subcorpus).[14] We report *percentage swing $d^\%$* and the results of two types of goodness of fit χ^2 test.

Overall, the perfect auxiliary declines in frequency by about 10% across the two subcorpora (Table 1, Total row). However, when the data for the different inflectional categories are considered, it becomes clear that not every perfect subtype behaves in the same manner. Past and infinitival subtypes fall by around 34% and 30% respectively, whereas the present (which is by far the most frequent of the subtypes) is stable. The figures for the *-ing* participle are small, and its fall is not statistically significant.

---

[12] Such stative examples are included in some corpus counts of the perfect (e.g. Biber *et al.* 1999: 463–7, who report GET as the most frequently occurring verb in this construction in British English conversation).

[13] This does exclude some genuine perfect examples. The results of Table 1 were checked against an alternative calculation which excluded only stative and ambiguous examples involving *got* for the present tense category, based on estimations from the random samples. This produced a very similar result, as few examples were clearly non-stative.

[14] These data, and further data reported in the paper, have been obtained from a revised version of the corpus prepared by the authors and others at the Survey of English Usage. The word counts used exclude 'ignored' material (material excluded from the structural analysis because it represents nonfluencies such as repetitions and reformulations).

Inflectional category	LLC		ICE-GB		Change in frequency		
	raw	pmw	raw	pmw	$d^\%$	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**	

Table 1. Frequencies of perfect auxiliaries in DCPSE, divided by inflectional category. Columns A and B represent goodness of fit χ^2 comparisons summarised in the text. Results marked 's' are significant at $p<0.05$.[15] Figures exclude certain instances of HAVE *got*: see discussion in text.

In order to be confident in the results 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 with the overall **set of perfect auxiliaries**.[16] This second comparison allows us to compare the *relative* change of each of the terms within this set, i.e. against a baseline of all perfect auxiliaries. Thus, whereas the slight percentage increase (0.75%) of the present subtype is not considered significant compared with the number of words (Column A), this subtype significantly increases its *share* of the set of perfect cases (Column B). Similarly, the table reveals that both past and infinitive forms decline to a greater extent than the overall trend would suggest. For the much less frequent *-ing* participle, neither result is significant.

Figure 4 displays changes in pmw frequencies for each form, from LLC (1960s data) to ICE-GB (1990s), with confidence intervals depicted by error bars.[17]
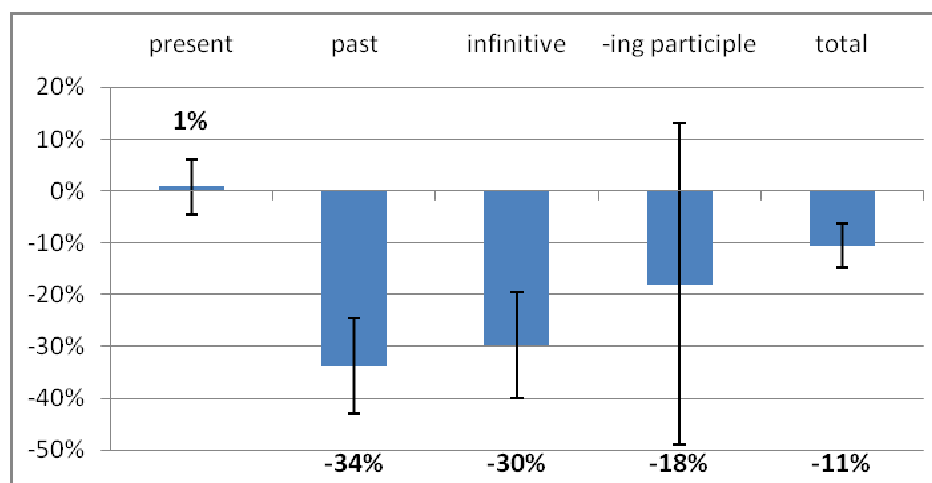

Figure 4. Changes in pmw frequencies (Table 1 '*d%*' column) with error bars representing 95% confidence intervals.[18]

[15] All significant results in this paper (marked 's') are evaluated at the p<0.05 level. For more on distinguishing different results by *effect size*, see the discussion in Section 2.2.3.

[16] To be more precise, 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 (Column A) or the perfect auxiliaries (B). The test evaluates whether the observed percentage change is significant (i.e. significantly *different from zero*).

[17] A confidence interval is a statistically estimated range of values within which the true value for a population can be assumed to fall with a certain level of confidence. These intervals were computed using Newcombe's Interval 10 for the difference between two proportions (Newcombe 1998), which is based on the Wilson score interval. This is a more precise method than traditional Gaussian error bars. See also Aarts *et al.* (this volume).

[18] This type of visualisation displays the size of the result (i.e. the column height) and our confidence in it. Another way of expressing the results for a given column, say 'past', is that we are 95% certain that the *past perfect* case falls by between 25% and 43%. Where an error bar does not cross the zero axis, the change can be said to be statistically significant (equivalent to a 2 × 2 test).

2.1.3   Discussion

For the present perfect, our results can be compared with those of Hundt and Smith (2009) for written British English in the LOB and FLOB corpora (comprised of printed material from 1961 and 1991 respectively). Note that Hundt and Smith's data were not manually reviewed so would include instances of stative HAVE *got*, although these are likely to be much less frequent in written than in spoken material (cf. Biber *et al*. 1999: 463–7).

First, we can observe that pmw frequencies of the present perfect are far higher in our spoken material from DCPSE, at around 6,000 pmw, than in the written corpora, at around 4,000 pmw (although there is variation among text categories or genres, a point to which we return later).[19] Second, neither study finds a significant change in pmw frequency from the 1960s to the 1990s (Hundt and Smith find a slight, non-significant decrease in frequency, while our study finds a very slight, non-significant increase). Hundt and Smith also looked at the present perfect in the written American corpora Brown and Frown, and again found no significant change in overall pmw frequency from the 1960s to the 1990s. Their data confirm that there is regional variation, with the present perfect in American English starting from a lower level of frequency and still at a significantly lower level than in British English. They therefore interpret their data as showing a pattern of regional variation which is stable over the time period.[20]

In our study we also considered the present perfect as a member of the set of all perfect auxiliaries. The significant result in Column B for the present perfect indicates that it differs significantly from the behaviour of the overall set, i.e. that the present perfect 'bucks the downward trend' and increases its share of the set. What this result means depends on our interpretation of why, and when, speakers choose to use the present perfect. Does *he has seen them* alternate with *he saw them* or *he sees them*? If alternation cannot be inferred in most cases, which we believe is likely, does the probability of use depend on context? We return to this question in Section 2.2 below.

Turning to the past and infinitival perfect categories, the observed declines raise further questions. One question concerning the infinitival perfect arises from the fact that the majority of instances in the corpus (88%) occur following a modal auxiliary. The modals themselves have declined in frequency by 6.4% (as a proportion of words) across the two subcorpora of DCPSE (Aarts *et al.*, forthcoming 2012); a recent decline has also been reported in studies of the modals based on other corpora (e.g. Leech *et al*. 2009).

This raises the question of whether the decline observed in the infinitival perfect is due simply to a decline in this major context of potential occurrence. This question is addressed in Bowie and Aarts (forthcoming 2011); a brief summary follows. FTFs were constructed to distinguish occurrences by their context. The proportions in which a perfect infinitive occurred in the two types of context (following a modal, and following *to*) were then obtained. The data revealed that the perfect infinitive declined significantly over this period in *both* types of context. Therefore the overall decline in frequency of the infinitival perfect cannot be attributed to a decline in the frequency of modal auxiliary usage.

A second question arises concerning the declines in both the past and infinitival perfect. Do they reflect a genuine change in usage, or could they be due to chance differences in *temporal orientation* in the texts sampled in the two subcorpora? That is, could speakers in the later subcorpus simply be talking less frequently about past time than those in the earlier subcorpus, due to chance sampling factors? In order to factor out such variation, we next consider subtypes of the perfect as members of a set of tensed, past-marked VPs.

---

[19] A higher frequency of the present perfect in spoken than in written English has also been reported in several other corpus-based studies without a diachronic dimension (compared in Schlüter 2006). Gries (2006) reports a similar finding for spoken and written data in the ICE-GB corpus (measuring frequency per number of verb forms rather than number of words, and measuring average frequency per file for each sub-register), but notes that this has to be qualified at the levels of register and sub-register: for example, the broadcast sub-register is an outlier with a particularly high frequency, and some written sub-registers have higher frequencies than some spoken sub-registers.

[20] This conclusion of course applies only to the standard regional varieties represented in the corpora.

2.2   The perfect in relation to tensed, past-marked VPs

Tensed, past-marked VPs (henceforth 'TPM VPs') were defined as tensed VPs marked for past either morphologically or with the perfect, or both. By 'marked for past' we mean formal marking, involving forms characteristically (though not of course invariably) used to indicate pastness or anteriority.

Four categories of TPM VPs were distinguished:

- past non-perfect (with 'past' indicating the morphological tense), e.g. *He **saw** them*
- present perfect, e.g. *He **has seen** them*
- past perfect, e.g. *He **had seen** them*
- modal auxiliary + perfect infinitive, e.g. *He **may have seen** them*

The majority (88%) of perfect infinitives appear with a preceding modal auxiliary. This category was included because modal auxiliaries are generally considered to be tensed forms; examples with both present and past tense modals were included in this category. Each category includes progressive and passive VPs: these represent separate dimensions of contrast.

Excluded from the set were present non-perfect VPs and non-tensed VPs. The exclusion of non-tensed VPs meant that both the *-ing*-participial perfect and the minority subset of the perfect infinitive occurring after *to* were excluded. While there are different possible ways of grouping VPs, it was decided to exclude non-tensed VPs because they offer at most only one choice for past marking (namely, the perfect).

Focusing on TPM VPs is important in factoring out variation between text samples that arises simply from differences in the degree of reference to past time. Note however that we do not go as far as to claim that the four categories of TPM VPs represent a *true alternation set*, where any member will alternate with any of the others in all instances.

2.2.1   Data retrieval

A set of FTFs was constructed to retrieve instances of the four categories of VP. As this method involves categories of VP, we employed queries for perfect auxiliaries which included only those instances where they occur in relation to VPs: either as daughter of VP, or preceding the VP within the same clause (e.g. in interrogatives where the first auxiliary precedes the subject). Figures 5 and 6 show the FTFs used to search for the present perfect and past perfect (with the addition of the relevant tense features). For the category modal auxiliary + perfect infinitive, we used the FTF shown in Figure 3 above (with both auxiliaries as daughters of VP) and a second FTF where the modal auxiliary occurs as a daughter of the clause and precedes a VP containing a perfect infinitive.
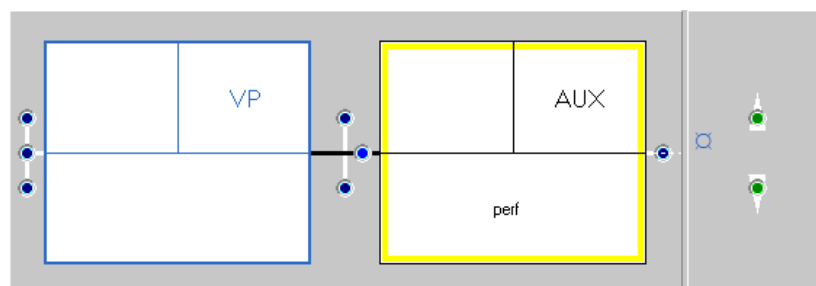

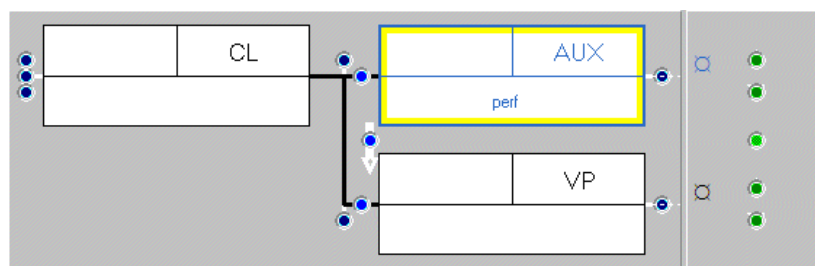Figure 5. FTF for perfect auxiliary as daughter of VP.

Figure 6. FTF for perfect auxiliary preceding the VP.

Unlike the simpler FTF queries described in 2.1, these searches excluded the minority of perfect auxiliaries (in the present or past tense) which occur in tag questions and short text units with no lexical verb, such as *has he*. As before, all instances parsed as perfect auxiliaries with *got* as past participle were manually excluded for the present perfect category; stative or ambiguous instances with *got* were excluded for the past perfect category (being included instead in the past non-perfect category).

The other past-marked category is the non-perfect. DCPSE marks past-tense VPs as 'past', which are easy to retrieve with a single-node FTF. Using negation it is possible to construct a single-node FTF which excludes VP nodes that are marked as perfect. However, to obtain the greatest precision, we opted for a more detailed approach.[21]

We employed an initial search for (i) VPs including a past auxiliary or main verb, (ii) clauses with a past tense auxiliary preceding the VP (or the 'predicate element', PREDEL), and (iii) clauses with a past-tense main verb acting as an operator, e.g. in an interrogative like *Was it her daughter* (these are labelled as V, parsed as daughters of the clause with no intervening VP node).[22] It was then necessary to subtract cases belonging to the other categories (past perfect, and past-tense modal + perfect infinitive).

To obtain figures for all VPs (for the baseline comparison), we combined the results of a search for VPs categorised as such (excluding 'empty' coordinating nodes) with a search for clauses with main-verb operators in either past or present tense (cf. (iii) above).

2.2.2   Results

We will consider the results in each of the four categories against two baselines: all VPs and only those VPs that are tensed and past-marked. The percentage changes in frequency per VP from LLC to ICE-GB are summarised by the graph in Figure 7. Percentage changes as a proportion of TPM VPs are shown in Figure 8.

Consider first the per-VP results in Figure 7. We observe a significant decline in frequency of around 12% for all TPM VPs from the 1960s to the 1990s. That is, as a proportion of the VPs in the corpus, 12% fewer VPs are tensed and past-marked in the later ICE-GB subcorpus of DCPSE than in the LLC subcorpus.

Against the VP baseline there is also a significant decline of a similar magnitude for the past non-perfect category. The categories past perfect and modal + (infinitival) perfect show significant declines of larger magnitude: around 35% (with a confidence interval of ±9%) and 27% (±11%), respectively. The present perfect appears stable. The results for the perfect categories are similar to those shown for pmw frequencies (in Table 1 and Figure 4).

---

[21] The figures turn out to be very similar in the latest revision of DCPSE, although cases in subcategory (iii) above would not be identified by the simpler approach.

[22] See Nelson *et al.* (2002) for further details of the grammatical analysis used in the corpus.
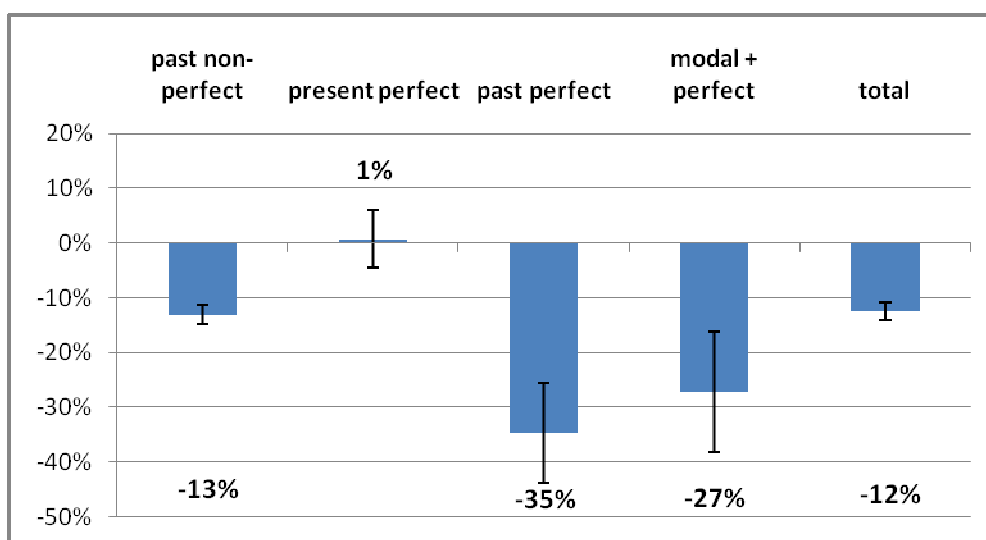
Figure 7. Changes in frequency of TPM VP categories as a proportion of all VPs.

Figure 7 indicates that the four subcategories seem to behave differently from one another over time. However, this variation could be explained in part by variation in the degree of 'pastness' of a text.

We next consider each category as a proportion of *tensed, past-marked VPs* rather than all VPs, and test whether any change in frequency differs from the overall behaviour of the set. As the set of TPM VPs comprises around 30% of all VPs in the corpus, we are narrowing our baseline for comparison considerably. Table 2 summarises the analysis, with percentage changes in the proportion of TPM VPs in each subtype illustrated in Figure 8.

This analysis obtains a somewhat different picture from before. The past perfect and modal + (infinitival) perfect types show significant decreases relative to the overall set, although these are of a smaller magnitude than those in Figure 7. The past non-perfect (which is by far the largest category, accounting for about 80% of all cases of TPM VPs) now appears to be stable when considered as a proportion of TPM VPs only. On the other hand, the present perfect, which appeared stable when compared to VP frequency, significantly increases as a proportion of the set. The changes observed are sizeable, as indicated by the percentage differences ($d^{\%}$). As a further measure of effect size, we cite ϕ' values obtained using Wallis' method to apply Cramér's ϕ to a goodness of fit test (see 2.2.3 below for discussion, and Appendix for details on this method).

TPM VP subtypes	LLC		ICE-GB		Change in frequency (per TPM VP)		
	raw	% of total	raw	% of total	$d^{\%}$	χ^2	ϕ'
past non-perfect	16,858	80.67%	13,451	80.16%	-0.64%	0.31	0.0032
present perfect	2,696	12.90%	2,488	14.83%	14.92%	**25.06 s**	0.0694
past perfect	782	3.74%	468	2.79%	-25.47%	**25.49 s**	0.1460
modal + perfect	561	2.68%	374	2.23%	-16.98%	**7.79 s**	0.0928
total TPM VPs	20,897	100.00%	16,781	100.00%	0.00%		

Table 2. Changes in frequency of subtypes as a proportion of the set of TPM VPs, including goodness of fit χ^2 for each subtype ('s' means significant at $p<0.05$) and goodness of fit ϕ' effect size.
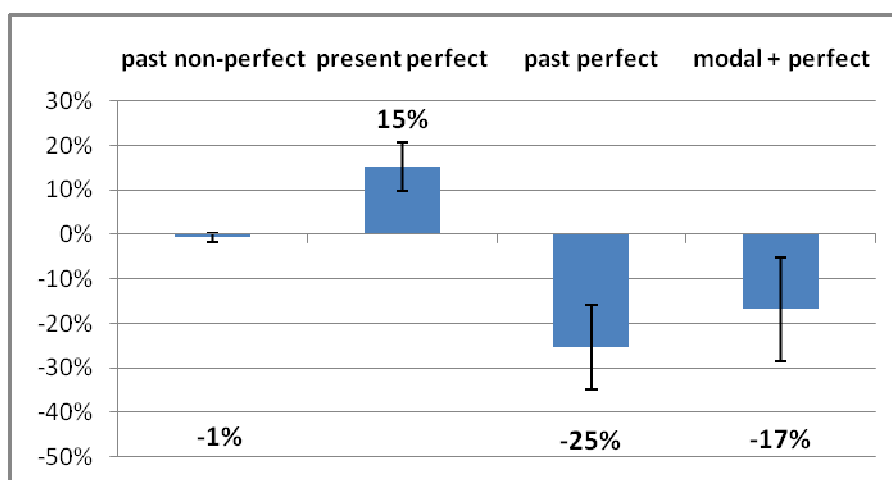
Figure 8. Changes in frequency of categories as a proportion of TPM VPs only.

How may we interpret these findings? On the face of it, use of the past non-perfect appears stable, while the present perfect is increasingly chosen at the expense of the past perfect and modal perfect subtypes. However, this explanation implies that subtypes freely alternate with each other, so that one subtype replaces another. Our earlier survey of the semantic distinctions between the subtypes (in the introduction to section 2) clarifies that this is not the case. The correct explanation likely lies elsewhere.

We turn to this problem of alternation in more detail in section 3. In the meantime, an alternative hypothesis suggests itself. It may be that the apparent 'stability' of the past non-perfect masks opposing trends of change: a decrease to the benefit of the present perfect, and an increase at the expense of the past perfect and modal + perfect. However, the only way to thoroughly investigate this possibility is to identify alternation patterns in each of the non-perfect past cases. Reviewing some 30,000 cases, or even a sizeable subsample, was unfortunately beyond the scope of our paper.

We may compare pairs of categories, however, to investigate how robust our findings are. We make pairwise comparisons of each of the perfect categories with the past non-perfect category, since each can be thought of as contrasting with that form in a different way (*he saw* vs *he has seen*, *he saw* vs *he had seen*, etc.). For example, we look at the proportions of the present perfect and past non-perfect within a set comprising those two categories only, and at the percentage changes in proportions over the two parts of the corpus.

The results for the pairwise comparisons, displayed in Figure 9, are similar to those obtained when the categories are considered as proportions of the group of four categories of TPM VPs. In pairings against the past non-perfect, we find a significant increase for the present perfect (around 13%),[23] and significant decreases for the past perfect (24%) and for modal + perfect (16%). The past non-perfect is by far the most frequent category, and therefore does not significantly vary in any of the three comparisons. In this case the precise choice of baseline does not make a big difference to the outcome.

---

[23] As proportions of present perfect vs past non-perfect are sometimes reported in the literature, we note our figures here for comparison. For the corpus as a whole, we find 14.61% present perfect and 85.39% past non-perfect. Across the earlier and later subcorpora, the percentage of present perfect increases from 13.79% to 15.61%.
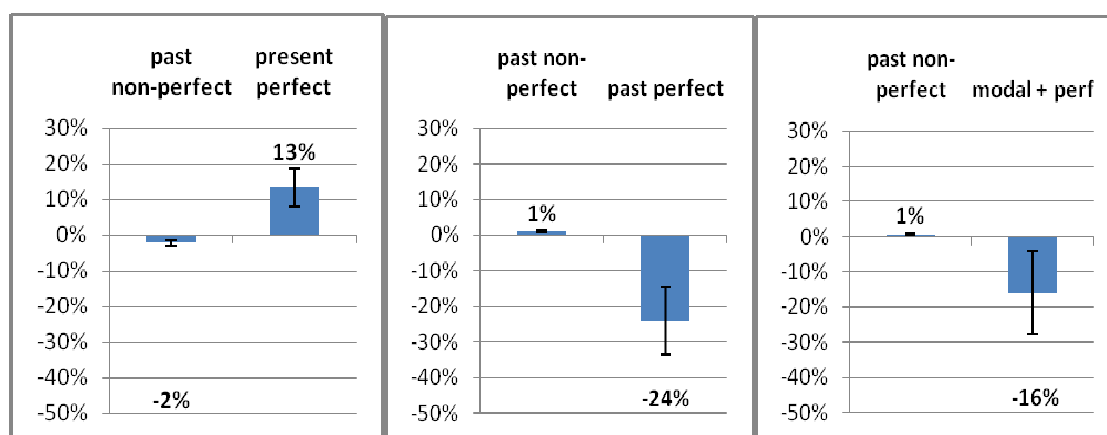
Figure 9. Changes in relative frequency of perfect categories against the past non-perfect.

One last issue is our decision to include modal cases, as modality is a dimension of meaning distinct from that of time reference. As a further test, we therefore remove the modal cases and compare the non-modal TPM VPs (past non-modal non-perfect VPs, present perfect VPs, and past perfect VPs). As shown in Figure 10, the findings for these categories are again similar to those for the four-way comparison: a significant decrease for the past perfect (approximately 25%), a significant increase for the present perfect (15%), and no significant change for the past non-perfect. The results are robust, regardless of which groupings are made for comparison.
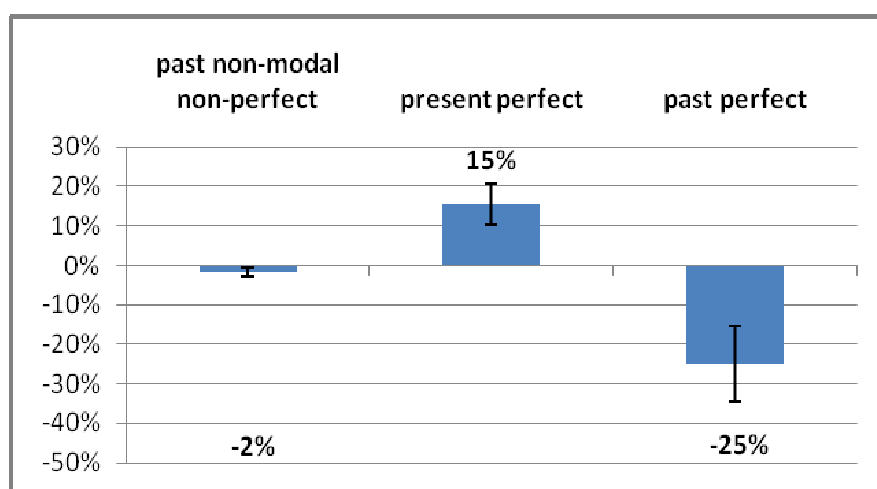
Figure 10. Changes in relative frequency of non-modal TPM VPs.

2.2.3 Discussion

Gries (2006) discusses a number of methodological issues concerning genre variation, taking the present perfect as a case study. He raises several important points regarding a tendency for researchers to over-report statistically significant differences due to minor differences in sampling or text type. In brief, with large amounts of data, any difference is likely to be statistically significant and it is important to look at measures of *effect size* as well as statistical significance. *Effect size* is a measure of how much the independent variable (time, in our case) impacts on, or correlates with, the dependent variable (e.g. the proportion of TPM VPs that are, say, past non-perfect). *Statistical significance* may be defined as the threshold at which the size of the effect is considered too large to be explicable by

mere chance.[24] The percentage difference $d^{\%}$ plotted in our graphs and Cramér's ϕ each measure effect size in different ways (for more information see the Appendix).

Gries also supports quoting frequencies as a proportion of a superordinate grammatical term (such as VP frequency). However, in his paper he does not go so far as to argue for considering a term relative to a set of its alternants.

Although DCPSE was constructed by grouping similar texts in the London–Lund Corpus and ICE-GB in a balanced manner as far as possible, some of the observed variation we are citing as being due to 'time' may be attributable to unavoidable sampling differences and genre variation. Our contention is that our focus on TPM VPs allows us to factor out a great deal of contextual variation that may arise as a result. Here it is not necessary to maintain that every instance of a TPM VP will alternate with every other subcategory. Rather, we merely claim that TPM VP frequency obtains a more reliable measure of the extent to which each text refers to past events than other measures such as VP or word frequency. We therefore consider the total set of TPM VPs as our baseline and consider changes in frequency of particular forms relative to this baseline.

The changes we find are both significant and sizeable. We obtain ϕ' scores of between 0.07 (present perfect) and 0.15 (past perfect) for cases of significant variation over time within the TPM VP set: these are small to medium sizes of effect, but are not insubstantial given that we have factored out other sources of variation.

We further showed the robustness of our results by testing different groupings of categories: similar results were obtained in (i) pairwise comparisons of each of the three perfect categories against the past non-perfect only, and (ii) a three-way comparison of the present perfect and past perfect against only non-modal past non-perfects.

When considered as proportions of past-marked forms, the past perfect and infinitival perfect categories each show significant and sizeable decreases, whether in larger groupings or in pairwise comparisons against the past non-perfect only. This evidence justifies the conclusion that the observed declines in these categories reflect genuine changes in usage, rather than being explained by an overall increase in the extent of reference to past time. These declines are further investigated in section 3.

As for the present perfect, the findings are more difficult to interpret and require further discussion. First, we can again compare the results for this category with those of Hundt and Smith (2009) for written British and American corpora. As mentioned in section 2.1, their study and ours agree in finding no significant change when figures for the present perfect are considered as frequencies per million words. However, the two studies obtain different results when the present perfect is considered as one of a group of past-marked forms. Hundt and Smith grouped the present perfect with the past tense only, and found that proportions of the present perfect remained stable. In our data, proportions of the present perfect increased significantly in all comparisons: when it was paired with the past tense only, and when it was grouped with a set of two or three other tensed, past-marked forms.

The correct interpretation of this observed increase is unclear. The uncertainty arises from the specialised pattern of use of the present perfect, involving an orientation towards *present* as well as past time. On the one hand, the present perfect does provide one option for locating situations in past time, and we have followed other studies in measuring its frequency against other past-marked forms more generally and/or the past non-perfect in particular (e.g. Elsness 1997; Hundt and Smith 2009). On the other hand, it is an option for marking past time that is likely to occur in broader contexts which are also oriented towards present time. It is therefore unclear whether the observed relative increase reflects a genuine increase in usage. It could instead be simply a result of the decline of TPM VPs in the corpus, and a concomitant increase in present-oriented contexts.

Does the present perfect behave more like the TPM VP set or an alternative tensed *present-marked* VP set? To answer this question we can compare the effect of using either baseline. Table 3a finds a non-significant result for the hypothesis that the present perfect changes as a proportion of

---

[24] In the case of large data sets, a small effect may be significant. Therefore to compare two different evaluations one should always compare effect sizes, not *p* values in significance tests.

present-marked VPs (including the present perfect). That is, we cannot distinguish present perfects from the total set. This result contrasts with the behaviour of the present perfect against TPM VPs (see Table 3b, expanded from Table 2).

present	LLC	ICE-GB	Total	present perfect goodness of fit
present non-perfect	33,131	32,114	65,245	$d^{\%}$ = -4.45 ± 5.13%
present perfect	2,696	2,488	5,184	ϕ' = 0.0227
TOTAL	35,827	34,602	70,429	χ^2 = 2.68 ns

Table 3a. Comparing present perfect cases against present-marked non-perfect VPs. The low χ^2 and ϕ' scores indicate a close correlation with the overall set.[25]

past	LLC	ICE-GB	Total	present perfect goodness of fit
other TPM VPs	18,201	14,293	32,494	$d^{\%}$ = +14.92 ± 5.47%
present perfect	2,696	2,488	5,184	ϕ' = 0.0694
TOTAL	20,897	16,781	37,678	χ^2 = **25.06 s**

Table 3b. Comparing present perfect cases against TPM VPs.

The simplest explanation of the behaviour of the present perfect in our data, therefore, is that it changes in proportion to what we might call the 'presentness of the text'.

This brings us back to the issue of genre variation, as temporal orientation is likely to be influenced by genre. For example, we might expect the DCPSE category of spontaneous commentary to be relatively present-oriented, and that of legal cross-examination to be relatively past-oriented – expectations which are borne out by a comparison of the proportions of present and past tenses in those categories. For future research, a detailed examination of genre variation in use of the present perfect in DCPSE would be one worthwhile way to further investigate the observed relative increase in frequency.[26] As part of this research, it would be useful to consider various features of the linguistic context which are likely to favour higher frequencies of the present perfect. For example, Elsness (1997) reports a tendency towards higher frequencies of this form in interrogative and negative constructions, and in clauses with first- or second-person subjects – features associated with more interactive contexts, and therefore likely to be highly variable in frequency across genres.

Co-occurrence with various kinds of temporal adjuncts could also be investigated, since this is an area where suggestions of current change have been made in the literature. However, a preliminary search for instances co-occurring with obvious specifications of past time finds no evidence of an increase in this pattern in DCPSE. We followed Elsness (2009b, a study of the present perfect in Australian and New Zealand English) in searching for the following convenient selection of items in co-occurrence with the present perfect: *ago*; *yesterday*; *last night/week/month/year/*day* (where **day* allows for *Monday*, etc.). We found only seven relevant examples, and in fact six of these occurred in the earlier subcorpus. Notably, in most examples a pause and/or hesitation signal intervenes between the present perfect and the temporal expression, as in the following example (where '<,>' and '<,,>' indicate short and long silent pauses respectively):

---

[25] Figures in Tables 3a and b are raw frequencies. The goodness of fit test compares the distribution across the two subcorpora of the present perfect with the distribution of the total category (present-marked or past-marked VPs). This tests whether the present perfect changes over time to the same degree as the 'total' category.

[26] For written English, Hundt and Smith (2009) report considerable variation in pmw frequency of the present perfect among the four broad categories of press, general prose, learned, and fiction. They also report some significant changes in pmw frequencies over time for particular genres (for British English, an increase in fiction and a decrease in the press category).

(8)   *I <,> think there's <,> very little we can do about it now because we**'ve sold** the pass <,,> uhm <u>seven or eight years ago</u> <,>* [DL-D01#311]

This is not, of course, a full investigation of temporal expressions (which would require considerable manual analysis), but it certainly suggests that an increase in co-occurrence with past time specification is unlikely to be found, and very unlikely to explain the observed proportional increase in frequency of the present perfect.

These issues concerning the present perfect must be left aside for future research. In the following section we focus on the perfect subtypes for which we found strong evidence of declining usage in DCPSE: the infinitival perfect and past perfect.

3   Infinitival perfect and past perfect: explaining their decline

In this section we discuss possible explanatory factors for the observed frequency declines in the infinitival perfect and past perfect, and report on their possibilities for alternation with non-perfect forms.

One likely explanatory factor behind the declines is an increasing tendency to simplify the verb phrase in this dimension of contrast (i.e. perfect versus non-perfect), particularly in contexts where the choice makes little or no difference to overall meaning. A second possible factor is American influence. As mentioned earlier, written corpus data presented by Elsness (1997) shows that these forms have fallen in frequency in American English from 1800 to the present day, and are less frequent in American English than in British English in the contemporary period. It has been suggested by Gorrell (1995) that the past perfect may be subject to ongoing decline in American English; he does not give diachronic evidence, but cites contemporary examples where a past non-perfect is used instead of an expected past perfect (cf. also Michaelis 1998: 269).

There is a difficulty with our suggested explanation: these two pressures, simplification and American influence, might also be expected to apply to the present perfect (which is also less frequent in American than in British English), yet it is not declining in frequency in our data. These contrasting trends need explanation.

One point is that low-frequency forms or usages may be likely to be less stable due to the distinctions that they convey being lost over generations. Leech *et al.* (2009: 90, cf. also 269–70), comparing frequency changes in individual modal verbs and in particular uses of modal verbs, are led to suggest 'a general principle that "losers lose out" (the infrequent suffer loss more than the frequent do)'. The idea arises from arguments by functional and cognitive linguists that greater usage strengthens cognitive representations (e.g. Evans and Green 2006: 114, 118–20). In our data the past and infinitival perfect are much lower in frequency than the present perfect (see Table 1). According to this hypothesis the distinctions conveyed by the past perfect and infinitival perfect will tend to be lost as they are increasingly replaced by non-perfect forms (and perhaps to a more limited extent also by the present perfect).

The specialised pattern of use of the present perfect, involving an orientation towards present as well as past time, may also be implicated in the contrasting trends. As we noted earlier, it has a distinct place in the 'ecological system' for expressing temporal relations. Thus there may be circumstances where it is uniquely appropriate so that its general use is being positively reinforced.

Returning to the infinitival perfect and past perfect, we have suggested that these forms may be subject to an increasing tendency to simplify the verb phrase, particularly in contexts where the choice between perfect and non-perfect forms has a minimal effect on meaning. The question is whether such contexts can be identified. To address this we consider each form in turn. We first discuss the infinitival perfect more briefly, summarising work presented elsewhere, and then turn in more detail to the past perfect.

3.1   Infinitival perfect

In the case of the infinitival perfect, Bowie and Aarts (forthcoming 2011) identify a number of instances in the corpus where the substitution of a closely similar non-perfect variant would make little or no difference to the overall meaning in context. Consider (9) and (10):

(9)   *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]

(10)   *he's forty odd I would **have thought*** [DI-B49 #53]

In (9), the second perfect seems superfluous, as no further anteriority is involved (compare *your Party would have had to **do** something*).[27] In (10), compared with a possible *he's forty odd I would **think***, the perfect merely reinforces the tentativeness with which the opinion is offered (the speaker evidently still holds the same opinion).[28] There is thus some leeway for the use of closely similar non-perfect structures instead of structures with a perfect infinitive; however, this leeway seems to apply only in a restricted set of instances.

In other instances, simple omission of the perfect produces a clear temporal contrast, so the only possible non-perfect alternatives would have to involve rather different structures. Consider the example *well I may **have missed** something* [DL-D01#0284]. Omission of the perfect (*well I may **miss** something*) changes the temporal location of 'missing something' from past to future. However, instead of using a structure with a modal auxiliary followed by a perfect infinitive, a speaker could express the modal meaning outside the verbal group and employ a simple past tense (e.g. *well <u>perhaps</u> I **missed** something*; *well it <u>may</u> be that I **missed** something*; *well it's <u>possible</u> I **missed** something*). Indeed, depending on the example and its interpretation in context, a present perfect (*well <u>perhaps</u> I've **missed** something*, etc.) may be substitutable, although by default one would expect the present perfect to be less often appropriate than the simple past (given the much higher frequency of the latter as a past-referring form). It is therefore difficult to specify a determinate set of alternating forms. A broader-ranging investigation of the expression of modality in combination with past time reference would be needed to fully identify speakers' changing choices over time.

3.2   Past perfect

It is somewhat easier to identify contexts of potential alternation between past perfect and past non-perfect in otherwise identical structures, as we discuss below. The past perfect is the more specialised form, expressing more differentiated meanings than the past non-perfect; it can therefore replace a past non-perfect only where these more differentiated meanings are appropriate, while replacement in the other direction (i.e. of a past perfect by a past non-perfect) means that the more differentiated meanings are not expressed in the verbal group (though they may be expressed elsewhere in context), with results which vary in their degree of acceptability.

The possibility of alternation is influenced by a number of factors, including for example the aspectual type of situation referred to (e.g. stative vs dynamic, durative vs punctual, bounded vs non-bounded), the expression elsewhere in context of the temporal relationships involved, and different conceptualisations of temporal relationships (see e.g. Declerck 1991, 2006). Given the multiple factors

---

[27] Such 'superfluous' multiple uses of the perfect have been condemned by prescriptive grammarians from the eighteenth century onwards (Molencki 2003).

[28] The tentativeness effect may come about because locating an opinion in past time suggests a readiness to revise it (cf. the discussion by Declerck and Reed 2005 of 'suspended factuality' in examples like *Oh I thought you didn't like ice cream*, uttered when confronted with possible counterevidence). Most instances of such formulae in the corpus are similar to (10) in that the perfect seems to merely add to the tentativeness expressed and so could readily be omitted. 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.

involved, the often graded nature of the acceptability of alternants, and uncertainty about implicit meaning distinctions, it would be difficult and time-consuming to identify with objectivity all and only those instances in the corpus where alternation is in fact possible, and so to provide exact quantification of changing patterns of choice in such contexts: among the 25,000-odd instances of non-modal, non-perfect past tense VPs, only a tiny minority are likely to allow alternation with a past perfect. We are confronted with the problem of identifying 'non-occurrence' of the past perfect: instances where the form with the more differentiated meaning could have been used but was not.

We therefore focus on (i) considering the contexts in which the past perfect occurs (using manual analysis of the 1,260 instances), and (ii) obtaining quantitative data comparing the frequencies of past perfect and (non-modal) past non-perfect VPs in relevant structural contexts which can be specified in automatic searches.[29] The following sections discuss alternation in several specific contexts – temporal clauses, backshifting contexts, and modal remoteness contexts – before considering potential alternation beyond these contexts.

3.2.1   Temporal clauses

One potential context of alternation noted in descriptive grammars (e.g. Quirk *et al.* 1985: 196; Huddleston and Pullum *et al.* 2002: 147) is that of subordinate clauses introduced by temporal expressions such as *after* or *as soon as*, which often suffice to make the temporal relations clear. The following corpus examples are closely similar, but contrast in their use of past perfect (11) versus simple past (12):

(11)   *I met him a lot <u>after</u> my father **had died** ...* [DI-A09 #4]
(12)   *They sent one to my mother <u>after</u> she **died** or something* [DI-B03 #277]

Although *when* expresses simultaneity, there are instances where this can be loosely interpreted as close succession, making alternation possible. Compare, for example:

(13)   *I went and did a bit of Christmas shopping <u>when</u> **I'd finished*** [DL-B19 #524]
(14)   *and you took something from a side table and you ate it and <u>when</u> you **finished** you went away and there was no formality* [DL-B03 #221]

Only a small proportion (around 6%) of all instances of the past perfect in the corpus occur in temporal clauses, though the proportion is significantly higher in LLC (7%) than in ICE-GB (4%). Most are introduced by *after* or *when*, while the other forms used are *before*, *until*, *till*, *as soon as*, *since*, *once*, and *no sooner*.

Substitution of a past non-perfect does seem possible in many instances, but by no means all. In examples like (15), with *after* and an adverbial of duration, a past non-perfect seems unlikely although perhaps not impossible; while in many examples with *when*, like (16), substitution of a past non-perfect would change the temporal relationship between the two situations referred to.

(15)   *we just found out <u>after</u> we**'d been** there <u>a few days</u> that this system existed*
[DL-B30 #433]

(16)   *<u>When</u> **I'd done** it all I looked at it* [DI-B47 #322]

When comparison is made of (non-modal) past non-perfect VPs and past perfect VPs in clauses introduced by one of the set of temporal expressions listed above (those found in the past perfect examples), the proportion of past perfect VPs is found to drop significantly from LLC (at around 8%) to ICE-GB (approximately 4%), as shown in Table 4 (left).

---

[29] Modal VPs are excluded because the perfect/non-perfect contrast in that context involves the perfect infinitive. In the FTF searches for (ii), the two categories of VP were specified by using VP nodes with the relevant features ('past, perfect, not-modal' and 'past, not-perfect, not-modal' respectively).

However, most of the examples with the past non-perfect involve *when*, whose meaning of simultaneity most often rules out replacement with a past perfect (e.g. *Well how old were you when the tonsils were taken out* [DI-A10 #125]). When the same comparison is made with *when* omitted from the set of introducing forms, the proportion of past perfect VPs is again found to drop significantly from LLC to ICE-GB (from approximately 18% to 9%, Table 4, right). The evidence does therefore suggest an increasing tendency to simplify in the temporal clause context, although it can only be taken as suggestive, as this approach is not based on strictly delimited alternating contexts.[30]

	all cases				excluding *when*			
	LLC		**ICE-GB**		**LLC**		**ICE-GB**	
past	raw	%	raw	%	raw	%	raw	%
non-perfect	561	91.82%	439	95.85%	143	82.18%	113	91.13%
perfect	50	8.18%	19	4.15%	31	17.82%	11	8.87%
TOTAL	611	100.00%	458	100.00%	174	100.00%	124	100.00%

Table 4. Decline of past perfect usage in temporal clause contexts: left, all cases; right, excluding *when* to avoid references to simultaneous events. Results are significant (goodness of fit χ^2 at $p<0.05$).

3.2.2 *Backshifting*

Another context of potential alternation involves the special backshifted use of the past perfect. Backshifting to a past perfect can occur in contexts such as indirect reported speech or thought when the reported 'original' involves past time which would be expressed with a past non-perfect or present perfect (while backshifting to past non-perfect can occur with an original present non-perfect). In many instances which report an original with simple past tense, the backshifted past perfect is not necessary to ensure the correct temporal interpretation, so a tendency towards simplification is likely to favour the non-backshifted simple past tense (Huddleston and Pullum *et al.* 2002: 158; cf. also Sakita 2002). Example (17) has a backshifted past perfect while (18) to (20) each have a non-backshifted simple past tense.

(17) *Oh actually Dad asked me if Sarah **had phoned** me on Sunday* [DI-B19 #0122]
(18) *And you said you first **got** in touch with A S Consultants before Christmas* [DI-H01 #95]
(19) *Can you remember how you felt when you heard that she **died*** [DI-E06 #83]
(20) *Oh you said you **put** on a stone and a half* [DI-B30 #74]

Around 15% of instances of the past perfect in the corpus occur in obvious backshifting contexts, that is in subordinate clauses where there is a superordinate (higher-level) predicate of speech or thought involving either a past tense form or a past temporal reference.[31] The proportion in ICE-GB is higher than in LLC, but not significantly so.

Corpus searches were used to compare the numbers of past perfect and (non-modal) past non-perfect VPs occurring in clauses functioning as direct objects with higher past tense VPs. This was chosen as a set of structurally identifiable backshifting contexts. Scrutiny of a random sample confirmed that the great majority of these examples do involve higher predicates of speech or thought. However, the search excludes certain relevant contexts, notably those where the higher predicate lacks past tense form but has past temporal reference (e.g. *I had a letter from him saying that he'd been in*

[30] Note that, even with *after*, not all instances of the past non-perfect allow substitution of a past perfect. For example, in *I was in the navy five days after I was eighteen* [DI-B23#242] we could not substitute *after I **had been** eighteen* (though *after I **had turned** eighteen* would be possible).

[31] Backshifting is not restricted to these obvious contexts (cf. Huddleston and Pullum *et al.* 2002: 154). However, as the boundaries of its use in subordinate clauses are unclear, we have identified the obvious cases. There is also a use in main clauses to represent a character's thoughts or internal speech, but this literary device can be assumed to be rare in spoken material.

to see Brendan _{DL-A07#361}; such examples comprised a large minority of 18% of the past perfect examples in backshifting contexts). The comparison of clausal direct object structures found no significant difference in proportions of past perfect between LLC and ICE-GB. We cannot therefore conclude that the past perfect is declining in backshifting contexts.[32]

It should be mentioned that there is also potential for a backshifted past perfect to alternate with a non-backshifted *present* perfect (e.g. *She said he had arrived* vs *She said he has arrived*). This alternation is more constrained than that between past perfect and past non-perfect, because a non-backshifted present tense indicates that the original reference to present time is still applicable at the time of the report (e.g. *She said he has arrived* would not be used if he had arrived and then left again, as his arrival would no longer have current relevance). It can further be noted that instances of present perfect VPs in the structure mentioned above (clausal direct objects with higher past tense VPs) are very few by comparison with non-modal past non-perfect VPs (34 compared with 1,071 instances).

3.2.3   Modal remoteness

Another particular usage of the past perfect is to express the combination of modal remoteness and 'pastness'.[33] The modal remoteness use of the past tense conveys *remoteness from reality* (counterfactuality or unlikeliness, depending on the construction). Where this meaning is combined with past time, the past perfect is the expected form (e.g. Huddleston and Pullum *et al.* 2002: 149–50). One context for this use is the remote conditional clause. Example (21) is a case of the expected past perfect usage. However, sometimes instances occur where a simple past is used although the time reference is past. Gorrell (1995) cites a number of examples for American English (cf. also Quirk *et al.* 1985: 1012; Molencki 2000: 324). Examples (22) and (23) are two of a handful of instances found in DCPSE. Example (23) involves a past progressive where a past perfect progressive might have been expected.

(21)   *it would've been nice if I **had known*** _[DI-B13#85]
(22)   *I was bored. Well I would have been if I **did** the last year.* [i.e. the last year of study in a particular subject] _[DI-B40#16–17]
(23)   *And going back a couple of years to your playing days I don't think you would've been happy if your forwards **were palming** out* _[DI-F02#79]

In the corpus, around 10% of instances of the past perfect occur in modal remoteness contexts. The conditional clause introduced by *if* is the most common context. When comparison is made of past perfect VPs versus (non-modal) past non-perfect VPs in non-interrogative *if*-clauses,[34] the proportion of past perfect VPs is found to be higher in LLC (18%) than in ICE-GB (12%), but the difference is *not* statistically significant.

This kind of evidence is only suggestive, as there are of course many instances where alternation is not possible – for example where a past non-perfect VP is used to express past time in an open (non-remote) conditional (*we used to have fires lit in our room for us if we **were** ill in bed* _[DL-B13 #230]), or modal remoteness plus present or future time reference in a remote conditional (*how would you live if you **became** a full-time student* _[DL-A01 #596]). However, only 8 possible examples (3 in LLC and 5 in ICE-GB) of past non-perfect constructions with the relevant meaning were found in a search of the most likely contexts, namely conditional constructions where the superordinate clause contains a modal auxiliary with or without a perfect infinitive.

---

[32] As with temporal clauses, this experiment is not premised on a strict identification of possible alternation contexts. For example, many occurrences of the past non-perfect involve backshifting of an 'original' present tense and so could not be replaced with a past perfect.

[33] The 'pastness' expressed generally involves past time, or anteriority to the present, but can also involve anteriority to a future reference point.

[34] Interrogative instances, where *if* is equivalent to 'whether', were excluded.

There is thus little evidence in DCPSE of increasing use of the past non-perfect to express the combination of modal remoteness and pastness. However, evidence from other sources suggests that this is an area of ongoing change in English.

Here is an example from a recent UK newspaper article about an accident that happened abroad, quoting the comments of the coroner:

(24) *David Morris, sitting in Dunstable, Bedfordshire, said: 'The balcony came completely away from the wall and crashed down. She had absolutely no chance. I rather suspect the awning on which she was climbing up may have been more substantial if it **was constructed** here.'* [*Metro*, 25 November 2010]

There are several noteworthy features about this extract. First, notice the use of the present tense form of the modal *may* where *might* would be the expected form in the apodosis of this conditional construction.[35] Secondly, in the protasis the past tense is used, rather than the expected past perfect (*if it **had been constructed** here*). Further examples of an unexpected past non-perfect can be found using simple search strings such as '*would have * if*' in large unparsed corpora such as the *British National Corpus* (BNC) and (for American English) the *Corpus of Contemporary American English* (COCA):[36]

(25) *I mean last time I was at an opera was erm I had these tickets now erm I took my sister and a friend of hers. Do you know what the total price of the tickets would have been if I **had** to pay for this, one hundred and ten pounds. Mm and there's no way I would have paid that.* [BNC, HYY 132–134, S_meeting, 1993]

(26) *Well we'd have finished if we **sat** down when you did.* [i.e. 'finished our tea'] [BNC, KSS 4552, S_conv, 1992]

(27) *It must have been seventy below zero outside; Jezrael wondered how much worse it would have been if Mars **were** at aphelion.* [BNC FPO 599, W_fict_prose, 1993]

(28) *…and I went to school to become a mechanical engineer, which is what he would have done if he **had** the opportunity to go to school.* [COCA:2009:SPOK NPR_TalkNation]

(29) *None of this would have happened if you **didn't** do that with the knife.* [COCA:2006:FIC IowaRev]

(30) *Mudear would have laughed if she **knew** that her husband had ever considered her safety.* [COCA:1993 FIC Bk:UglyWays]

For many speakers a past perfect in these examples would have been more natural, yet the same speakers would probably not stumble over the examples as they stand. This might indicate a degree of insecurity about this syntactic configuration. Note also the use of irrealis (or 'past subjunctive') *were* rather than *was* in (27), another possible variant when the verb is BE.

In many contexts there is cliticisation of *had* onto the preceding subject and concomitant phonetic reduction, which may be conducive to the disappearance of this auxiliary over time. (Around 40% of instances of the past perfect auxiliary in DCPSE occur in the cliticised form *'d*.) In some cases of assimilation the /d/-sound may disappear, as in the following example from a BBC news broadcast reporting on young women being accosted in the street by strangers in a car.

(31) *They make the right decision and walk away. If they**'d got** into the car police were prepared to move in.* [BBC News at Ten, 24 November 2010]

---

[35] Huddleston and Pullum *et al.* (2002: 202–3) note such uses of *may* as dialectal variants which are becoming increasingly common.

[36] The BNC is available online via several interfaces: as the BYU-BNC (see Davies 2004–); through the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium (www.natcorp.ox.ac.uk); or as BNCweb (http://bncweb.info). For COCA online, see Davies (2008-).

A phonetician who listened to the clip for us had no hesitation in transcribing *if they'd got* rather than *if they got*: phonetically *'d* in *had* and the initial sound in *got* are assimilated such that they become a long /g/-sound, held for too long to be the single /g/ of *got* preceded by *they* (Michael Ashby, personal communication). Apart from this subtle feature of timing, it then sounds as though *had* has been dropped altogether.[37] Furthermore, as with many other verbs, the past participle form of GET is identical with the past tense form, which also reduces the distinctness of the past perfect from the past non-perfect.[38]

   Another context for the modal remoteness use is within the complement of **wish**, as in the corpus example *I wish he **had come** round yesterday* [DI-B34#28]. There are no examples in DCPSE where a past non-perfect is used in this context with reference to past time, but a few examples have been found in other sources; note the use of irrealis *were* in (35) (from American English).

(32)   *I wish you **were** at the shack with me last night* [BNC, KP6 2307, conversation, 1993]

(33)   *I wish I **got** this fish in butter* [BNC, KC6 2046, conversation, 1992]

(34)   *You're sitting, waiting, wishing you **left** your house earlier. Why not use this spare time to sharpen your mind with More Brain Training on Nintendo DS? ...* [advert on tube train, London, 2010]

(35)   *The saga of Joe Montana goes on, and while everybody involved says he wishes it **were** over yesterday, it may still be several days or even weeks.* [American syndicated sports story, cited by Gorrell 1995: 26]

The modal remoteness use presents a further complication, namely the occurrence of the so-called 'double perfect' or variants thereof, as seen in the following examples of remote conditionals:

(36)   *Though I think if you**'d have asked** me I wouldn't have said I did* [DI-E026#128]

(37)   *What, if the comedians **had have been** an hour late, would you have booked another comedian?* [BNC, J92 535, meeting, 1993]

(38)   *If more **would have gone** back then the strike would have come to an end quicker* [BNC, HMM 226, oral history interview, 1986]

(39)   ***Had** he **have lost** this frame, it would have been all over for him.* [BBC commentator, during a televised snooker context; cited by Declerck 2006: 27]

In this use, there seems to be a counteracting tendency towards a more complex, expanded verbal group (with infinitival *have* variously preceded by *had*, *'d*, or sometimes *would*). Such forms are often considered non-standard but may now be spreading into standard usage (cf. Huddleston and Pullum *et al.* 2002: 151); they are rare in DCPSE (with only one example each in LLC and ICE-GB) but seem to be quite common in the BNC (where they also occur in clausal complements of **wish**, e.g. *I wish they **had have attended**, the Tories* [JJD 219, meeting, 1994]).[39] Example (39) is interesting in that it occurs in an example with an inverted conditional clause (a past non-perfect variant would not be possible here: **Did he lose this frame, it would have been all over for him*).

   The occurrence of these 'expanded' verbal groups seems to be due to the frequent association of the perfect infinitive with the meaning of modal remoteness, and a pull towards parallelism

---

[37] This example is unusual in that the apodosis does not directly express modal remoteness: police were in fact prepared to move in, and it can be inferred that they would have done so if the condition expressed in the protasis had been fulfilled. Note also that, while the preceding sentence uses the present tense to refer to a past situation presently unfolding on screen, a present tense would not be possible in a non-modally remote version of the sentence under consideration (i.e. we could not have *If they get into the car police are prepared to move in*, because the women have already walked away).

[38] Elsness (1997) discusses phonetic reduction of the auxiliary, and frequent syncretism between past tense and past participle, as factors tending towards loss of the present perfect; we do not find a decline for the present perfect in our data, and have suggested there are countervailing factors where the present perfect is concerned.

[39] Numerous British English examples of *'d/had + have/'ve* in conditional clauses are also cited by Declerck and Reed (2001: 193–4), mainly from the spoken component of the Cobuild Corpus of English.

between the protasis and apodosis of the remote conditional (Denison 1998; cf. also Molencki 2000). In fact, such usages are not completely new in English, being attested from as early as the 15th century (Denison 1993: 356–7; Molencki 2000). Molencki (2000) suggests later prescriptivism as a cause of reduced variation among alternative forms in past 'counterfactual' contexts in Early Modern English.

In summary, examples from both British and American sources support the suggestion of Leech (2004: 128) that the area of meaning involving modal remoteness and past time is one of instability and possible ongoing change in present-day English. Further investigation of the multiple variant forms, and of their distribution in different varieties of English, would be worthwhile.

3.2.4 Other contexts

All the examples of alternation discussed so far occur in quite specific subordinate clause contexts (involving meanings of temporal relationship or modal remoteness). However, the possibility of alternation between past perfect and past non-perfect extends beyond those specific contexts. In this section we consider examples in main and relative clauses.

In the following two examples from American English (the first involving a main clause and the second a relative clause), a past non-perfect is used where one might have expected the writer to use a past perfect in order to express a temporal relationship of anteriority.

(40) *He looked rather pitiful. In years past he **was** a tiger, a ruthless intimidator who tied even the slickest lawyers in knots.*
[John Grisham, *The Pelican Brief*, 1992, cited by Gorrell 1995: 27]

(41) *Serb mortar crews targeted a pick-up soccer game ... on Tuesday ... An hour after two shells slammed into the crowd of spectators, the soccer ball remained on the parking lot near Sarajevo airport where the game **was played**.*
[Austin *American Statesman*, June 2, 1993, cited by Michaelis 1998: 269]

In (40) the temporal relationships are made clear by the expression *in years past*, while in (41) they have to be inferred from the context without the support of any overt temporal expression.

In British English also, there is sometimes the possibility of alternation between past perfect and past non-perfect in more general contexts such as main and relative clauses, as shown by examples from DCPSE.[40] Compare (42) and (43), each involving a coordination of two main clauses which refer to a sequence of two events. In each, a past perfect is used to refer to the first event. However, in (42) a past perfect is also used to refer to the second event, while in (43) a simple past is used although the second event, like the first, involves time anterior to a time which has been referred to in preceding context.[41]

(42) *Sarah <u>had</u> <u>left</u> the kitchen roll in the sitting room and Chicken **had** methodically **shredded** it* [DI-B45#43]

(43) *I'<u>d</u> <u>phoned</u> the day <u>before</u> and the doctor **said** it wasn't necessary* [DL-H01#167]

Subtle differences in conceptualisation can also lead to contrasting choices, as suggested by the following main clause examples from DCPSE:

(44) *Oh I **hadn't realized** that* [DL-B36#59]

---

[40] Declerck (2006) has noted this possibility for British English, commenting that 'the temporal relation which exists between two situations in the real world is not necessarily expressed by the verb forms of the clauses reporting these situations' (p. 446).

[41] As the past-in-past temporal relationship has been established in the first clause in (43), it may be that the speaker has, in the second clause, shifted to treating this as the primary past domain under discussion. That is the kind of analysis proposed by Declerck (1991), who sees the simplification as going hand-in-hand with a semantic or conceptual shift.

(45) *I **didn't** realize that* [DL-B34#225]

Each of these occurred as a response to something the speaker had just been told. The being told and the realising happened just a moment ago, a time which can be thought of as 'just then' or as 'now'. The past perfect example seems to imply 'before you told me', and the past tense example 'before now'. There seems to be just a subtle difference in meaning between the two when considered in context. (This perhaps also involves the type of situation, or Aktionsart, with the past perfect suggesting a dynamic 'achievement' interpretation of REALISE and the past tense suggesting a stative interpretation.)

In the following relative clause examples from DCPSE, a simple past is used to refer to a situation that is anterior to the situation of the superordinate clause, and so could readily be replaced by a past perfect:

(46) *I sort of felt that there was peace in a corner of my mind in which I always **felt** there was turmoil.* [The first 'felt' refers to the time after the speaker's mother had died.] [DI-E06#89]
(47) *so I just took uh some of the tablets you **gave** me and it cleared up within two days* [Context indicates the giving of the tablets preceded their taking by as much as a year.] [DI-A20#226]

The main clause and relative clause contexts are too general to allow easy retrieval and quantification of instances of a past non-perfect where a past perfect could be substituted. However, a quantitative comparison of different clause types provides suggestive evidence.

Almost two thirds (64%) of instances of the past perfect in the corpus occur in subordinate (or 'dependent') clauses, and the proportion does not differ significantly between LLC and ICE-GB (this is much higher than the proportion of all clauses which are dependent, 43% overall). Thus 36% of instances of the past perfect occur in main clauses. The proportion found in relative clauses is quite high at 18%. Taken together, main and relative clauses therefore account for more than half of the occurrences of the past perfect in the data.

When comparison of past perfect VPs and (non-modal) past non-perfect VPs is made within clause types, the proportion of past perfect VPs is found to decline significantly from LLC to ICE-GB in main clauses as well as in subordinate clauses; and also in relative clauses taken as a separate group of subordinate clauses. Non-relative subordinate clauses do not show a significant decline as a group, but we saw above (section 3.2.1) that a significant decline is found in the specific context of temporal clauses.[42] These comparisons suggest that there may be an increasing tendency to replace the past perfect with the past non-perfect in main and relative clauses, as well as in temporal clauses.

4.   Conclusion

In this paper we have presented corpus data on recent trends in the perfect construction in spoken British English, covering the period from the 1960s to the 1990s. We considered three main subtypes of this construction: present perfect, past perfect, and infinitival perfect. In order to investigate change in usage over time, we compared subtype frequency against baselines of (i) words, (ii) VPs, and (iii) the set of tensed, past-marked ('TPM') VPs. This set consists of tensed VPs marked for past either inflectionally or with the perfect, or both.

The first method found stability in the present perfect, but significant decreases in the past perfect and infinitival perfect. The second method obtained similar results, along with a decline in the past non-perfect. The third method again found decreases in the past perfect and infinitival perfect, but also an increase in the present perfect.

---

[42] In DCPSE, non-relative dependent clauses are assigned the features 'subordinate' or 'zero subordinate' depending on the presence or absence of an overt subordinator. To search for more specific types of dependent clause, more complex FTFs are required.

The largest category of TPM VPs in DCPSE is the past non-perfect, which is highly stable over time as a proportion of this group of VPs, accounting for four out of five instances. In this analysis the other types are effectively compared against this majority case. The relative decline in usage for the two categories of past perfect and infinitival perfect therefore appears to be robust, and to reflect genuine changes in usage. Two explanatory factors behind these observed falls were suggested: American influence and simplification.

The situation with the present perfect appears to be more complex. Its relative rise in frequency compared to TPM VPs contrasts with its relative stability compared to all VPs. However, it is unclear whether this relative rise reflects a genuine change in usage patterns. If the present perfect is more readily used within broader contexts which are primarily oriented towards the present, then the observed relative rise may be merely a consequence of the decline of TPM VPs in the corpus, and an associated increase in present-oriented contexts. In a further investigation we considered the behaviour of the present perfect as a proportion of the set of tensed, *present*-marked VPs. The result of this comparison is non-significant: the behaviour of the present perfect cannot be distinguished from that of the total set. Thus the simplest explanation of our data is that the present perfect changes in proportion to the 'presentness' of the corpus texts in DCPSE, while actual usage patterns stay stable over time. This hypothesis is consistent with its higher frequency and specialised meaning (involving reference to present as well as past time). Detailed investigation of genre variation in use of the present perfect could shed further light on this issue.

Ideally we would wish to be in a position to evaluate variation within a closed set of choices. This would allow us to claim that the relevant forms were alternating with each other and therefore any observed rise of one form must be at the expense of an alternate form. However, the set of TPM VPs includes cases which are unlikely to alternate, which alternate with only one of the other forms, or whose patterns of alternation are simply unclear. We can only make a weaker claim that the size of this set is likely to be an effective means for estimating the 'pastness' of any particular text (i.e. the degree to which the text tends to refer to the past).

In order to consider alternation, one must evaluate a contrast in specific contexts. The problem for this study is that constraints on usage appear to be primarily semantic. Even with a parsed corpus such as DCPSE, the identification of contexts where true alternation can occur is non-trivial.

Nonetheless we examined a series of contexts in which the past perfect might alternate with the past non-perfect. There is suggestive evidence from DCPSE for an increasing tendency to choose the past non-perfect in main clauses, relative clauses, and temporal clauses. There appears to be little evidence in DCPSE of such change in the special uses of backshifting and modal remoteness plus past time, although evidence from other sources suggests that the latter is an area of ongoing change. These studies lend plausibility to the idea of an overall trend towards simplification and loss of distinctions over time affecting the use of this grammatical category. American influence may be implicated in this trend. The simplification tendency may be further encouraged by the frequent occurrence of reduced phonetic forms, which can be subject to assimilation and virtual loss in some contexts.

This tendency towards simplification, however, does not seem to be part of a general simplifying trend in the verb phrase (in the sense of 'verbal group'). There is ample evidence for a recent increase in the frequency of the progressive construction, both in DCPSE (Aarts *et al.* 2010) and in other corpora (Mair and Leech 2006). The simplification trend therefore appears to apply to the perfect/non-perfect dimension of contrast only. Furthermore, it seems to apply only to the less frequent subtypes of the perfect (the past perfect and infinitival perfect), and not to the much more frequent present perfect, perhaps because less frequent forms are in general more likely to suffer loss.

Appendix: applying Cramér's φ to compare a subset distribution against its superset

Cramér's φ (Aarts *et al.*, this volume, Appendix 2) is a standard measure of effect size for $r \times c$ contingency tables which measures the degree by which the table is skewed.

A 2×2 table where all cells are equal, or values in either rows or columns are constant, obtains φ = 0. A table where the cells form the identity matrix [[1, 0], [0, 1]], or some multiple thereof, returns a value of φ = 1. The formula is very 'well-behaved' mathematically. It can be proved that it behaves in strict proportion to a perturbation of the flat matrix along the diagonal. Cramér's formula can be expressed as

$$\text{(E1)} \qquad \phi = \sqrt{\frac{\chi^2}{(k-1)N}}$$

where N is the total number of cases in the table and, for an arbitrary $r \times c$ table (rows × columns), k is the smaller of r and c.

This measure is a 'contingency correlation' score, similar to other types of correlation score, such as Pearson's r^2, used for variables expressing continuous number values. Cramér's φ is rarely used in linguistics, whereas many studies cite percentage swing (see Aarts *et al.,* this volume). There are, however, a number of known problems with percentage swing as a measure of change, not least that $d^{\%}$ can range over any value and the values obtained are difficult to compare, because each is defined as change relative to an initial value (which is itself likely to vary).

Can φ be reformulated as a 'goodness of fit' measure (like $d^{\%}$) to measure variation of one column against another baseline column? This would give us a standardised measure of 'effect size' which would make results easier to compare, and may even be extended to any number of values – unlike $d^{\%}$. The answer is yes, but if we simply insert the result of a goodness of fit χ^2 sum into E1 the result can exceed 1. We have to adapt the formula.

Wallis (2010) proposes the method employed in this paper.[43] We will use the present perfect data and baseline of TPM VPs (Table 3b) to illustrate the approach.[44]

Example

	O pres. perf.	**baseline** TPM VPs	**E** scaled	**E'** *N/k*	**O''** O×E'/E	**O'** O''×*N*/ΣO''
LLC	2,696	20,897	2,875.15	2,592.00	2,430.49	2,412.06
ICE-GB	2,488	16,781	2,308.85	2,592.00	2,793.13	2,771.94
TOTAL Σ	*N* = 5,184	37,678	5,184.00	5,184.00	5,223.61	5,184.00

Table A1. A worked example of Wallis' goodness of fit φ' using data from Table 3b. φ' = χ²(O', E') = 0.0694.

Method

Let us refer to the adjusted distributions as E' and O'. The idea is that after recalibration, E' will be constant (flat), and O' will vary from it, so we need to transfer the variation to O'.

We require E'$_i$ ≡ N/k where N = ΣO. We then recalibrate the observed distribution O' in two stages. First, each term E$_i$ has been adjusted by multiplying by E'$_i$/E$_i$. We therefore repeat this scaling transformation for every term in the observed distribution, thus:

$$\text{(E2)} \qquad O''_i \equiv O_i \times E'_i/E_i.$$

---

[43] Wallis (2012) compares a number of methods for estimating this goodness of fit φ, and concludes that for many-celled tables the optimum method is likely to be a probabilistically-weighted estimate, ϕ_p.

[44] The computation is also available on a spreadsheet at www.ucl.ac.uk/english-usage/statspapers/2x2chisq.xls.

However, O" is not now guaranteed to sum to N (note that in Table A1, $N = 5{,}184$). We therefore rescale again to obtain O':

(E3) $O'_i \equiv O''_i \times N/\Sigma O''$.

We then calculate ϕ' from a goodness of fit $\chi^2(O', E')$ using equation (E1).

References

Aarts, Bas, Joanne Close, and Sean Wallis (2010). Recent changes in the use of the progressive construction in English. In: Bert Cappelle and Naoaki Wada (eds.), *Distinctions in English grammar, offered to Renaat Declerck*. Tokyo: Kaitakusha. 148–67.

Aarts, Bas, Joanne Close, and Sean Wallis (this volume). Choices over time: methodological issues in investigating current change.

Aarts, Bas, Gerald Nelson, and Sean Wallis (1998). Using Fuzzy Tree Fragments to   explore English grammar. *English Today* 14: 3, 52–56.

Aarts, Bas, Sean Wallis, and Jill Bowie (forthcoming 2012). Profiling the English verb phrase over time: modal patterns. In: Merja Kytö and Irma Taavitsainen (eds.), *Corpus linguistics and the history of English*. Cambridge: Cambridge University Press.

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 Bas Aarts (forthcoming 2011). Change in the English infinitival perfect construction. In: Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *Handbook on the history of English: rethinking approaches to the history of English*. Oxford: Oxford University Press.

Davies, Mark (2004–). BYU-BNC: The British National Corpus. Available online at http://corpus.byu.edu/bnc.

Davies, Mark (2008–). The Corpus of Contemporary American English (COCA): 410+ million words, 1990–present. Available online at http://www.americancorpus.org.

Declerck, Renaat (1991). *Tense in English: its structure and use in discourse*. London: Routledge.

Declerck, Renaat, in collaboration with Susan Reed and Bert Cappelle (2006). *The grammar of the English verb phrase,* Volume 1, *The grammar of the English tense system: a comprehensive analysis* (Topics in English Linguistics, 60.1). Berlin: Mouton de Gruyter.

Declerck, Renaat, and Reed, Susan (2001). *Conditionals: a comprehensive empirical analysis* (Topics in English Linguistics, 37). Berlin: Mouton de Gruyter.

Declerck, Renaat, and Reed, Susan (2005). What is modal about *I thought that …*? *English Language and Linguistics* 9:2, 311–32.

Denison, David (1993). *English historical syntax: verbal constructions.* London: Longman.

Denison, David (1998). Syntax. In: Suzanne Romaine (ed.), *The Cambridge history of the English language*. Vol. IV: *1776–1997*. Cambridge: Cambridge University Press, 92–329.

Elsness, Johan (1997). *The perfect and the preterite in contemporary and earlier English* (Topics in English Linguistics, 21). Berlin; New York: Mouton de Gruyter.

Elsness, Johan (2009a). The present perfect and the preterite. In: Günter Rohdenburg and Julia Schlüter (eds.), *One language, two grammars?: differences between British and American English* (Studies in English Language). Cambridge: Cambridge University Press, 228–45.

Elsness, Johan (2009b). The perfect and the preterite in Australian and New Zealand English. In: Pam Peters, Peter Collins and Adam Smith (eds.), *Comparative studies in Australian and New Zealand English: grammar and beyond*. Amsterdam; Philadelphia: Benjamins, 89–114.

Engel, Dulcie M., and Marie-Eve A. Ritz (2000). The use of the present perfect in Australian English. *Australian Journal of Linguistics*, 20:2, 119–40.

Evans, Vyvyan, and Melanie Green (2006). *Cognitive linguistics: an introduction*. Edinburgh: Edinburgh University Press.

Fischer, Olga, and Wim van der Wurff (2006). Syntax. In: Richard Hogg and David Denison (eds.), *A history of the English language*. Cambridge: Cambridge University Press, 109–98.

Gorrell, Robert (1995). The future of past tenses. *English Today*, 11:4, 25–7.

Gries, Stefan Th. (2006). Exploring variability within and between corpora: some methodological considerations. *Corpora* 1:2, 109–51.

Heine, Bernd, and Tania Kuteva (2006). *The changing languages of Europe*. Oxford : Oxford University Press.

Huddleston, Rodney, and Geoffrey K. Pullum *et al.* (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Hughes, Arthur, Peter Trudgill, and Dominic Watt (2005). *English accents and dialects: an introduction to social and regional varieties of English in the British Isles*. 4th edn. London: Hodder Education.

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 (2004) *Meaning and the English Verb*. 3rd edition. Harlow, UK: Longman.

Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith (2009). *Change in contemporary English: a grammatical study*. Cambridge: Cambridge University Press.

Mair, Christian, and Geoffrey Leech (2006). Current changes in English syntax. In: Bas Aarts and April McMahon (eds.), *The handbook of English linguistics*, 318–42. Malden, MA; Oxford: Blackwell.

Michaelis, Laura A. (1998). *Aspectual grammar and past-time reference* (Routledge Studies in Germanic Linguistics, 4). London; New York: Routledge.

Miller, Jim (2004). Perfect and resultative constructions in spoken and non-standard English. In: Olga Fischer, Muriel Norde, and Harry Perridon (eds), *Up and down the cline: the nature of grammaticalization*. Amsterdam; Philadelphia: Benjamins, 229–46.`

Molencki, Rafał (2000). Parallelism vs. asymmetry: the case of English counterfactual conditionals. In: Olga Fischer, Anette Rosenbach and Dieter Stein (eds.), *Pathways of change: grammaticalization in English* (Studies in Language Companion Series, 53). Amsterdam: Benjamins, 311–28.

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, R. G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine* 17: 873–90.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985). *A comprehensive grammar of the English language*. London: Longman.

Sakita, Tomoko I. (2002). Discourse perspectives on tense choice in spoken-English reporting discourse. In Tom Güldemann and Manfred von Roncador (eds.), *Reported discourse: a meeting ground for different linguistic domains* (Typological Studies in Language, vol. 52). Amsterdam; Philadelphia: Benjamins, 173–98.

Sampson, Geoffrey (2002). Regional variation in the English verb qualifier system. *English Language and Linguistics* 6/1, 17–30.

Schlüter, Norbert (2006). How reliable are the results?: comparing corpus-based studies of the present perfect. *Zeitschrift für Anglistik und Amerikanistik*, 54:2, 135–48.

Sheskin, D. J. (1997). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, FL: CRC Press.

Siemund, Peter (2004). Substrate, superstrate and universal: perfect constructions in Irish English. In: Bernd Kortmann (ed.), *Dialectology meets typology*. Berlin: Mouton de Gruyter, 401–34.

Svartvik, Jan (ed.) (1990). *The London-Lund Corpus of spoken English: description and research* (Lund Studies in English, 82). Lund: Lund University Press.

Tagliamonte, Sali A. (2000). The grammaticalization of the present perfect in English: tracks of change and continuity in a linguistic enclave. In: Olga Fischer, Anette Rosenbach and Dieter Stein (eds.), *Pathways of change: grammaticalization in English* (Studies in Language Companion Series, 53). Amsterdam: Benjamins, 329–54.

van Herk, Gerard (2008). The present perfect in early African American correspondence. *English World-Wide* 29:1, 45–69.

Walker, Jim (2008). The footballer's perfect: are footballers leading the way? In Eva Lavric, Gerhard Pisek, Andrew Skinner and Wolfgang Stadler (eds.), *The linguistics of football* (Language in Performance, 38). Tübingen: Narr, 295–303.

Wallis, Sean A. 2010. z-*squared: the origin and use of* χ^2. London: Survey of English Usage, UCL. www.ucl.ac.uk/english-usage/statspapers/z-squared.pdf

Wallis, Sean A. 2012. *Goodness of fit measures for discrete categorical data*. London: Survey of English Usage, UCL. www.ucl.ac.uk/english-usage/statspapers/gofmeasures.pdf