

# The English Perfect

Jill Bowie, Sean Wallis and Bas Aarts  
University College London

## 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 the* . The other subtypes are the past perfect (*She had seen the* ), the infinitival perfect (*She must have seen the* ), and the -ing-participial perfect (*Having seen the 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 The perfect in English

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

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).<sup>3</sup> 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 pre

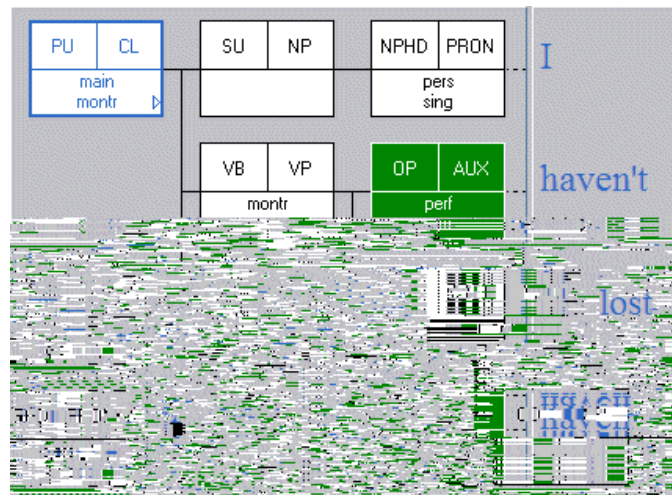


Figure 1. Tree diagram for the sentence *I haven't lost it.*<sup>6</sup>

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).<sup>7</sup> 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.

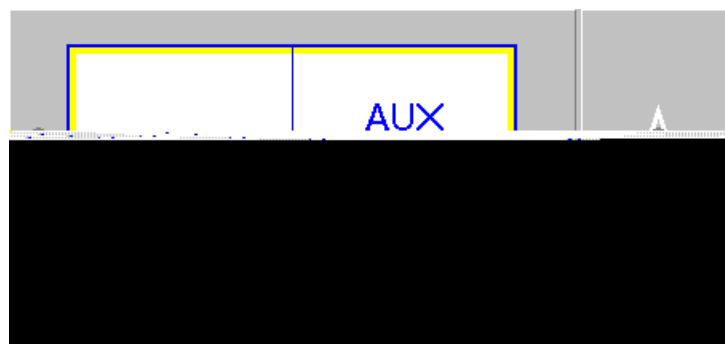


Figure 2. A simple FTF to search for a present perfect auxiliary.

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 have been*).<sup>8</sup>

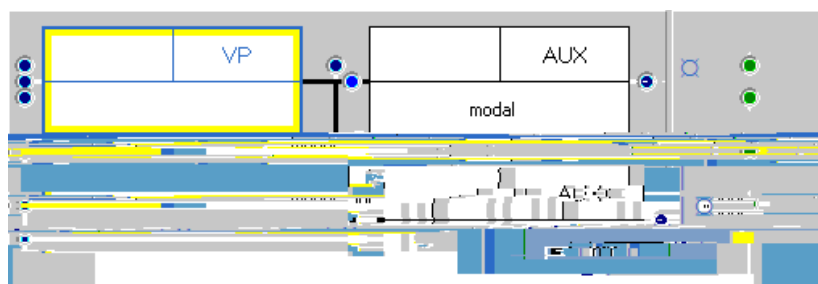


Figure 3. FTF for an infinitival perfect auxiliary following a modal auxiliary under a VP. The white arrow requires that the second auxiliary follows the first, but not necessarily in strict succession.

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.

#### type of perfect frequency end

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



1      e ec      e pe      on d

## 2      *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/7







dependence on opened perfect

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 the*
- present perfect, e.g. *He has seen the*
- past perfect, e.g. *He had seen the*
- modal auxiliary + perfect infinitive, e.g. *He may have seen the*

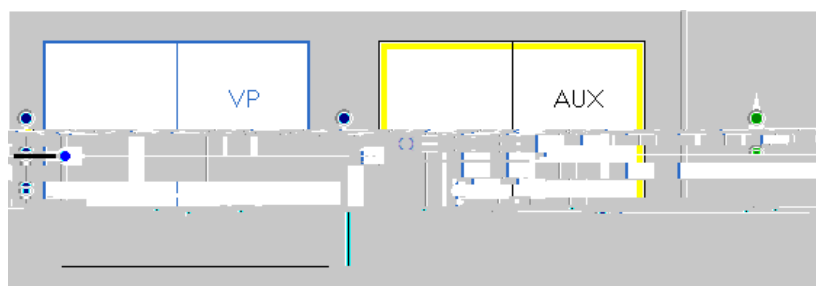
The majority (88%) of perfect infinitives appear with a preceding modal auxiliary. This category was included because modal auxiliaries are generally considered to be tensed forms; examples with both present and past tense modals were included in this category. Each category includes progressive and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

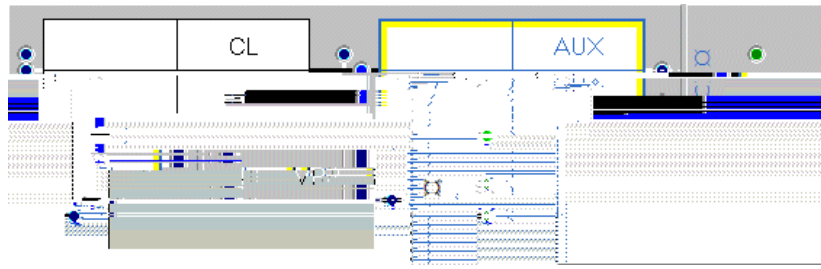
Excluded from the set were present non-perfect VPs and non-tensed VPs. The exclusion of non-tensed VPs meant that both the *-ing*-participial perfect and the minority subset of the perfect infinitive occurring after *to* were excluded. While there are different possible ways of grouping VPs, it was decided 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 tern tion set*, where any member will alternate with any of the others in all instances.

## 2.2 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.







past non-perfect present perfect past perfect modal perfect





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).

<b>p e en</b>	<b>LLC</b>	<b>C<sub>p</sub></b>	<b>Total</b>	<b>present perfect goodness of fit</b>
present non-perfect	33,131	32,114	65,245	$d = -4.45 \pm 5.13\%$
present perfect	2,696	2,488	5,184	$\phi' = 0.0227$
<b>AL</b>	<b>35,827</b>	<b>34,602</b>	<b>70,429</b>	$\chi^2 = 2.68$ ns

Table 3a. Comparing present perfect cases against present-marked non-perfect VPs. The low  $\chi^2$  and  $\phi'$  scores indicate a close correlation with the overall set.<sup>25</sup>

<b>p</b>	<b>LLC</b>	<b>C<sub>p</sub></b>	<b>Total</b>	<b>present perfect goodness of fit</b>
other TPM VPs	18,201	14,293	32,494	$d = +14.92 \pm 5.47\%$
present perfect	2,696	2,488	5,184	$\phi' = 0.0694$
<b>AL</b>	<b>20,897</b>	<b>16,781</b>	<b>37,678</b>	$\chi^2 =$

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









However, most of the examples with the past non-perfect involve *hen*, whose meaning of



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 company completely  
 was from the shed door She had soute y no chance I rather suspect  
 the thing on which she was sitting up y have been ore sustanti **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 *was* where *might* would be the expected form in the apodosis of this conditional construction.<sup>35</sup> 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 ‘*could have \*if*’ in large unparsed corpora such as the *British National Corpus* (BNC) and (for American English) the *Corpus of Contemporary American English* (COCA):<sup>36</sup>

- (25) *I went to the station to see her I had these tickets no more I took my sister  
 and friend of hers Do you know what the total price of the tickets could have been **if I  
 had to pay for this one hundred and ten pounds More and there is no way I could have  
 paid that** [BNC, HYY 132–134, S\_meeti]*









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 h







