Profiling the English Verb Phrase over Time: Modal Patterns[1]

Bas Aarts,
Sean Wallis
and Jill Bowie

University College London

1 Introduction

In recent work it has been demonstrated by various scholars that the use of modal verbs has changed in past decades (Krug 2000, Leech 2003, Smith 2003, and especially Leech *et al*. 2009). In earlier work Close and Aarts (2010) and Aarts, Close and Wallis (forthcoming) looked at the changing use of the modal and semi-modal auxiliaries in spoken English. While the results of these investigations are extremely valuable, they tell us very little about how modal verb usage has changed within the various modal verb phrase patterns. To our knowledge no one has ever looked at the frequencies of these various patterns, let alone how they have changed over time. In section 2 of this chapter we briefly introduce the *Diachronic Corpus of Present-Day Spoken English* (DCPSE; Aarts and Wallis 2006) which has been used as a database for the present research. In section 3 we look at previous research on changes in the use of modal verbs as individual items ('core modals') in written English. We will compare our findings on spoken English with those obtained by Leech (2003) and Leech *et al.* (2009) for written English in section 4. In section 5 we look at relative changes, while in section 6 we will present the results of the research we have done on modal verbs occurring in various patterns, and how modal usage in these patterns developed over time. In section 7 we discuss our results. The final section is the conclusion.

2 The database: the *Diachronic Corpus of Present-Day Spoken English*

DCPSE is a corpus of parsed (grammatically analysed) spoken English which contains approximately 885,000 words: 464,000 taken from the *London-Lund Corpus* (LLC) which was collected between the 1960s and 1970s (Svartvik 1990 and Nelson *et al.* 2002) and 421,000 words from the British component of the *International Corpus of English* (ICE-GB) dating from the early 1990s. The TOSCA/ICE phrase structure analysis is a highly detailed 'traditional' one based on Quirk *et al.* 1985, familiar to many linguists.

The corpus exploration software ICECUP 3.1 (the *International Corpus of English Corpus Utility Program*, Nelson *et al*. 2002[2]) was specifically developed for working with parsed corpora. This software manages the corpus, allows a range of queries to be performed and combined, and provides a number of overview tools (corpus map, lexicon and grammaticon) which allow data to be extracted for further analysis. A key idea of the software is that searching and obtaining numbers is not sufficient. Using a complex grammar for research requires that we are confident that our results are accurate – and mean what we *think* they mean.[3]

Trees consist of nodes representing part of speech tags, phrases and clauses, plus functional labels. Every word is given a node and crossing links are not permitted. For reasons of space we generally display these from left-to-right, with the text down the right hand side, although other orientations are possible. The tree diagram for the sentence *It's amazing, isn't it* (text code DI-B37, unit 206) is shown in Figure 1.

As can be seen, nodes contain three types of information: *functional information* (top left; e.g. subject, direct object, noun phrase head, etc.), *categorial information* (top right; e.g. pronoun, verb, clause, etc.), and optional additional *features* (bottom half; e.g. 'intrans' for intransitive (verb), 'sing' for singular, etc.).

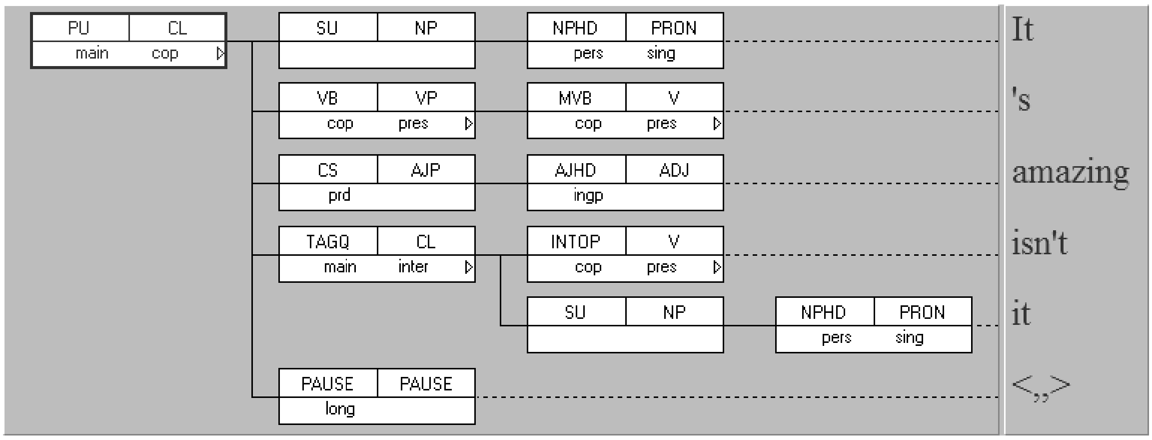

Figure 1
An example of a tree diagram in the DCPSE corpus, *It's amazing isn't it* $_{\text{DI-B37 \#206}}$.

PU = parsing unit, CL = clause, cop = copular, SU = subject, NP = noun phrase, NPHD = noun phrase head, PRON = pronoun, pers = personal, sing = singular, VB = verbal, VP = verb phrase, pres = present, MVB = main verb, V = verb, CS = subject complement, AJP = adjective phrase, prd = predicative, AJHD = adjective phrase head, ADJ = adjective, ingp = *–ing* participle, TAGQ = tag question, inter = interrogative, INTOP = inverted operator.

ICECUP permits users to construct grammatical queries called *Fuzzy Tree Fragments* (FTFs; Wallis *et al.* 2000; Nelson *et al.* 2002). FTFs can be composed of a single node, as in the example below which finds all modal auxiliary verbs in the corpus.

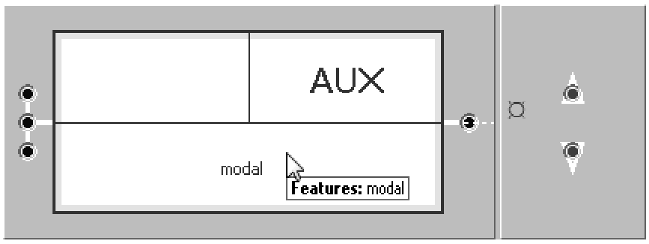

Figure 2
An example of a Fuzzy Tree Fragment which searches for all the modal verbs in the corpus.
AUX = auxiliary (form label), 'modal' = modal (feature label).

FTFs are, as the name suggests, *tree fragments*, involving patterns of nodes. The following FTF will find a clausal pattern involving a subject realised as a noun phrase, a verb phrase and direct object realised as a clause.

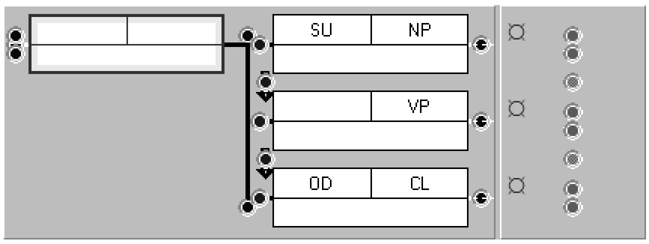

Figure 3
An example of a Fuzzy Tree Fragment which searches for a 'subject + verb phrase + direct object' pattern in the corpus.
SU = subject, NP = noun phrase, VP = verb phrase, OD = direct object, CL = clause.

Other FTFs, including some we will come to later, may involve more complex patterns, lexical wild cards, etc. The symbol '¤' in the FTFs above indicates an unspecified lexical element. Notice that in the FTF in Figure 3 the arrows between the nodes are black, indicating that the nodes in trees must follow each other immediately. Users can change these settings if they wish. FTFs provide an extremely intuitive user-driven search facility for grammar which can be made as simple or as complex as necessary.

3 Previous research on modal verbs in written English

Several scholars have reported that the core modals in written English have decreased in frequency (Krug 2000, Leech 2003, Smith 2003, and especially Leech *et al*. 2009). Figure 4 is taken from Leech *et al.* (2009) and shows the decline in the use of the core modals in written American and British English over the period 1961–1991/2. The data for British English are from the matching *LOB Corpus* (*Lancaster-Oslo-Bergen*) (1961) and its 1990s Freiburg counterpart *FLOB*, while the American data are from the matching one-million-word *Brown* (1961) and *Frown* (1992) corpora.

Note that references in this paper to ‘core modals’ include other forms based on the same modal auxiliary lemma, so *can* includes *can’t* and *cannot*, *will* includes *’ll* and *won’t*, etc.

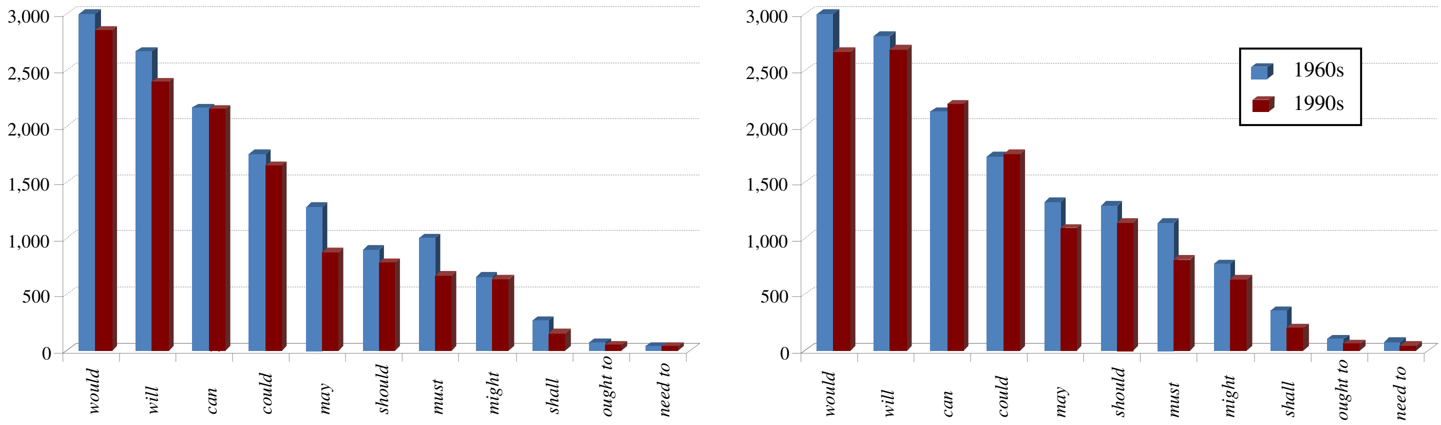

Figure 4
Frequency, normalised per million words, of modal auxiliaries in American English 1961–1992 (left) and British English, 1961–1991 (right). From Leech *et al.* (2009: 74), including *ought* (*to*) and *need* (*to*).

Not all modals are equally common, as Figure 4 reveals. The overall ranked order is similar for US and UK English, with *would* and *will* being the most frequent core modals, tailing off with *ought* (*to*) and *need* (*to*), which are classed as ‘semi-modals’. The latter will be excluded from the remainder of this discussion.

To measure change over time we will follow Leech *et al*. (2009) and employ the percentage difference of the normalised frequency, or ‘percentage swing’ for short, which we denote by $d^{\%}$. This is obtained from the following simple formula:

$$\textit{percentage swing } d^{\%} = (p_1 - p_2) / p_1.$$

Here p_1 and p_2 represent a proportion of a particular item, at two points, 1 and 2. In Table 1 and 2 p_1 and p_2 measure the proportion of words that are particular core modals (*would*, *will*, etc.) in the 1960s and 1990s respectively. We refer to this per-million-word change henceforth as *absolute* change. In Section 5 we employ the same formula to calculate change in the proportions of *core modals* that are particular modals, i.e. we measure change *relative* to the set of core modals. We will see that focusing on relative change has a number of benefits, in particular, it allows us to factor out and distinguish between multiple sources of variation.

To aid comparison of our results, we also carry out a z test to compare the swings for ‘statistical separability’, i.e. whether one observed change is significantly different from another (Wallis 2010). We will mark separable results in our tables with inequality signs (‘≠’).

Overall the results show a decline in usage for the core modals of around 10 percent in written BrE and AmE (the results are not statistically separable), shown in the final ‘Total’ row of Table 1. Many of the changes are individually significant (i.e. the change is 95% certain to be non-zero), and our analysis agrees with Leech *et al*. In addition, we can see that the changes in pmw density of *will*, *must* and *shall* are statistically different between US and UK data.

Table 1
Recent absolute changes in per-million-word (pmw) density of individual lexical core modals, adapted from Leech *et al.* (2009: 283; we excluded *ought* and *need* and added the statistical comparison of the Am/Br changes). Changes recorded in the $d^{\%}$ column are statistically significant (at $p<0.05$) unless otherwise stated ('ns').[4] Inequality signs ('≠') in the central column indicate where reported $d^{\%}$ trends for AmE and BrE significantly *differ*.[5]

Modal	**Brown (1961)**		**Frown (1992)**		**Change %**		**LOB (1961)**		**FLOB (1991)**		**Change %**
	raw	**pmw**	**raw**	**pmw**	$d^{\%}$		**raw**	**pmw**	**raw**	**pmw**	$d^{\%}$
would	3,053	3,012	2,868	2,854	**-5.2%**		3,032	3,011	2,682	2,660	**-11.7%**
will	2,702	2,666	2,402	2,390	**-10.3%**	≠	2,822	2,803	2,708	2,686	-4.2% ns
can	2,193	2,164	2,160	2,150	-0.7% ns		2,147	2,132	2,213	2,195	+2.9% ns
could	1,776	1,752	1,655	1,647	-6.0% ns		1,741	1,729	1,767	1,753	+1.4% ns
may	1,298	1,281	878	874	**-31.8%**		1,333	1,324	1,100	1,091	**-17.6%**
should	910	898	787	783	**-12.8%**		1,301	1,292	1,148	1,139	**-11.9%**
must	1,018	1,004	668	665	**-33.8%**	≠	1,147	1,139	814	807	**-29.1%**
might	665	656	635	632	-3.7% ns		779	774	640	635	**-17.9%**
shall	267	263	150	149	**-43.3%**	≠	355	353	200	198	**-43.7%**
Total	**13,882**	**13,696**	**12,203**	**12,144**	**-11.3%**		**14,657**	**14,557**	**13,272**	**13,165**	**-9.6%**

4 Modal verb frequencies in spoken English

Are changes in core modal frequency in spoken British English similar to those found in its written counterpart? To answer this question we compare the results obtained by Leech *et al.* (2009) with those found in the spoken British English corpus, DCPSE. Table 2 below shows the corresponding frequencies for the core modals in DCPSE, and the rates of change between the LLC and ICE-GB subcorpora, measured against the total number of words.[6] The written British English data is included for comparison.

Table 2
Recent changes in per-million-word density of individual lexical core modals in written and spoken British English.

Modal	**LOB (1961)**		**FLOB (1991)**		**Change %**		**LLC (1960s)**		**ICE-GB (1990s)**		**Change %**
	raw	**pmw**	**raw**	**pmw**	$d^{\%}$		**raw**	**pmw**	**raw**	**pmw**	$d^{\%}$
would	3,032	3,011	2,682	2,660	**-11.7%**		1,929	4,331	1,627	4,025	**-7.1%**
will	2,822	2,803	2,708	2,686	-4.2% ns	≠	1,493	3,352	1,528	3,780	**+12.8%**
can	2,147	2,132	2,213	2,195	+2.9% ns		1,618	3,633	1,568	3,879	+6.8% ns
could	1,741	1,729	1,767	1,753	+1.4% ns	≠	853	1,915	682	1,687	**-11.9%**
may	1,333	1,324	1,100	1,091	**-17.6%**		390	876	213	527	**-39.8%**
should	1,301	1,292	1,148	1,139	**-11.9%**		495	1,111	386	955	**-14.1%**
must	1,147	1,139	814	807	**-29.1%**	≠	467	1,049	194	480	**-54.2%**
might	779	774	640	635	**-17.9%**	≠	302	678	288	712	+5.1% ns
shall	355	353	200	198	**-43.7%**		193	433	90	223	**-48.6%**
Total	**14,657**	**14,557**	**13,272**	**13,165**	**-9.6%**		**7,740**	**17,379**	**6,576**	**16,267**	**-6.4%**

We can visualise change graphically by plotting $d^{\%}$ for each modal verb, including confidence intervals.[7] Where the confidence interval crosses the x axis (0%) the change is significant ('significant' means that the observed change is distinct from zero).

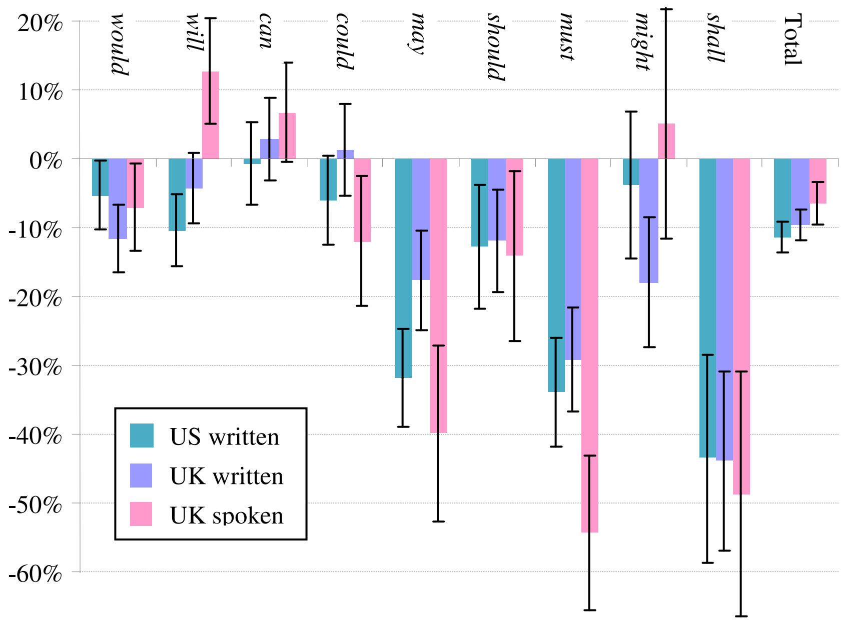

Figure 5
Change in pmw frequency from 1961 to 1991 for the core modals in US and UK corpora, expressed as a percentage of 1961 data.

Overall the spoken data seems to differ more markedly from the comparable BrE written data than the written BrE data differs from the AmE data.

There are approximately 20% more core modals per million words in the spoken data than the written, supporting the claim that modals are more frequent in spoken than in written language as reported by several scholars (see e.g. Biber *et al.* 1999: 486). This higher usage is holding up. There is a smaller observed fall in modal usage in a spoken context in our data (-6.4 percent vs. -9.6 percent).[8] At first sight this would seem surprising as one might have expected that the decline in modal usage in spoken English would be greater than the decline in written English.

Table 2 shows both whether a change is significant and whether the two observed changes are statistically separable. Not every numerically different result is sufficiently different from another for us to report that the difference would be repeatable were the corpus to be re-sampled 19 times out of 20. It turns out, for instance, that despite their apparent difference the two trends for all modals are *not* statistically separable.

Whereas the overall trend is still downwards, modal *will* has actually increased in frequency. We found a significant increase of 12.8 percent for *will/'ll/won't* compared with Leech *et al.*'s finding of a non-significant decrease of 4.2 percent (see also Aarts, Close and Wallis forthcoming). There is also a significantly greater observed fall in the use of *must/mustn't* (-54.2 percent vs. -29.1 percent) and trends for both *could* and *might* are similarly statistically separable.

5 Measuring relative change

So far we have considered change measured as frequency counts normalised per million words. This is a measure of 'absolute' change. Individual core modals are simply compared with a baseline of the number of words in the corpus. However, as we have seen, not all modals behave in the same way, and we often wish to distinguish variation of individual modals within an overall trend of all core modals.

One reason for doing this is simply that, as we have already seen with spoken data, some types of 'text' have a higher density of core modals than others. Therefore we would wish to factor out variation between corpus samples that arises due to differing density of core modals in texts. Similarly, it is axiomatic that, at the level of the utterance, modals do not appear randomly and independently, but rather occur in patterns, an issue we return to in this paper. This is a bigger problem with smaller samples dependent on fewer texts.

The type of analysis outlined above can be readily modified to consider how individual core modals change over time *relative to the overall set of modals*. To do this we replace the per-million-word column with the simple proportion of core modals in each category. Thus in the 1961 US Brown data, the lexical item *would* represents 3,053 out of 13,882 core modals (0.2199 or 22%). In this case the baseline is the total number of core modals, not the total number of words. The result of all three contrasts is shown in Table 3.

Table 3
Examining changes in the *relative* proportion of usage of individual lexical core modals.

Modal	Brown	Frown	$d^{\%}$		LOB	FLOB	$d^{\%}$		LLC	ICE-GB	$d^{\%}$
would	0.2199	0.2350	**+6.87%**		0.2069	0.2021	-2.31% ns		0.2492	0.2474	-0.73% ns
will	0.1946	0.1968	+1.13% ns	≠	0.1925	0.2040	**+5.97%**	≠	0.1929	0.2324	**+20.46%**
can	0.1580	0.1770	**+12.05%**		0.1465	0.1667	**+13.83%**		0.2090	0.2384	**+14.06%**
could	0.1279	0.1356	+6.01% ns	≠	0.1188	0.1331	**+12.08%**	≠	0.1102	0.1037	-5.89% ns
may	0.0935	0.0719	**-23.05%**		0.0909	0.0829	**-8.87%**	≠	0.0504	0.0324	**-35.72%**
should	0.0656	0.0645	-1.62% ns		0.0888	0.0865	-2.55% ns		0.0640	0.0587	-8.22% ns
must	0.0733	0.0547	**-25.35%**	≠	0.0783	0.0613	**-21.63%**	≠	0.0603	0.0295	**-51.11%**
might	0.0479	0.0520	+8.63% ns		0.0531	0.0482	-9.27% ns	≠	0.0390	0.0438	+12.24% ns
shall	0.0192	0.0123	**-36.09%**		0.0242	0.0151	**-37.78%**		0.0249	0.0137	**-45.11%**

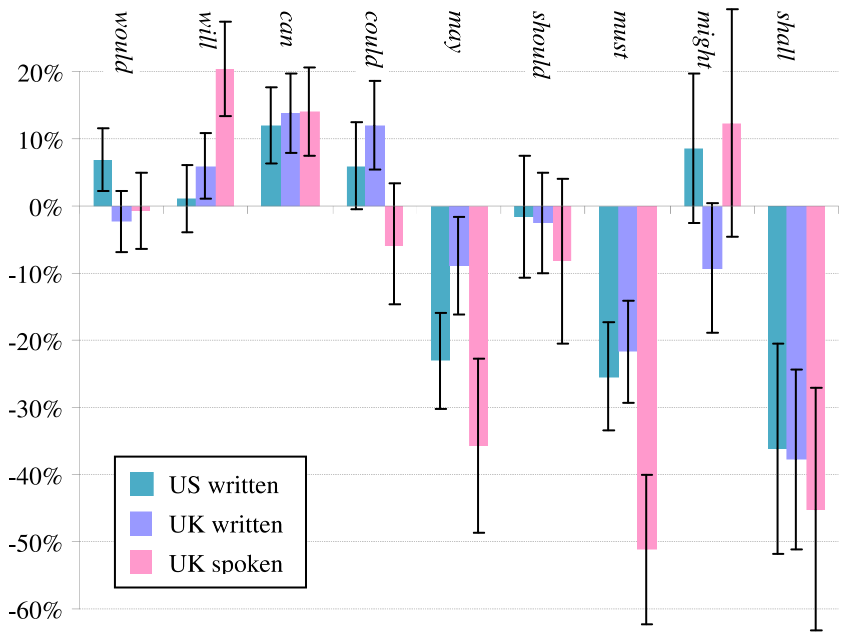

Figure 6
Recent change in the proportion of individual core modals in US and UK corpora, expressed as a percentage of 1961 data.

The picture obtained from an examination of relative change (Figure 6) differs notably from that for absolute change (Figure 5). The most obvious example concerns *can*, which, in all three corpus comparisons, increases as a proportion of core modals. When compared to the number of words, however, *can* appears static.

We can also see that *will* is increasing in relative terms for both UK spoken and written English. This tends to suggest that spoken English is 'leading the way' in promoting this change. Similarly, the falls in usage of *may* and *must* are significantly greater in the spoken data. There may be a question of transatlantic influence in the fall in usage of *must*, as the US written data is also falling significantly faster than the UK written data. Finally the trends for *could* and *might* also differ from each other, but they are not individually significant, with the exception of a rise in *could* in UK written English.

We could continue improving our analysis by focusing on semantically constrained sets of alternate forms, such as *can* vs. *may* or *will* vs. *shall*. We leave this to other papers (see, e.g. Aarts, Close and Wallis forthcoming). Thus far we have measured change relative to the baseline of 'all core modals', which is an improvement on 'all words', but does not represent a semantically bounded set. It is clearly not possible to replace, say, *will* with *might* without altering the meaning of an utterance. We cannot therefore state with certainty that *will* is replacing *shall* in Figure 6 above.[9] Nonetheless our data is highly suggestive of this trend.

6 Modal patterns in spoken English

In this section we will explore how modal verb usage has changed from a different perspective. We will do so by investigating the rates of change in the eight patterns in which modal verbs occur, listed below.

A: modal + main verb/VP
B: modal + perfect HAVE + main verb/VP
C: modal + progressive BE + main verb/VP
D: modal + passive BE + main verb/VP
E: modal + progressive BE + passive BE + main verb/VP
F: modal + perfect HAVE + progressive BE + main verb/VP
G: modal + perfect HAVE + passive BE + main verb/VP
H: modal + perfect HAVE + progressive BE + passive BE + main verb/VP

Within each of these patterns we recognise a number of sub-patterns, namely declarative and interrogative patterns, with or without intervening elements.

The result of this investigation is a rather complex experiment. Attempting to perform a series of FTF searches with ICECUP III and classifying them 'by hand' is likely to be too difficult in practice to be trustworthy. We need to accurately categorise each modal pattern into each type A1–A11, B1–B10 etc. We also need to be sure that each instance is considered to match each pattern once only.

This is where the new ICECUP IV software (Wallis 2008) comes in. This software allows us to design experiments within the platform. In our case we create a new classification 'variable' consisting of a series of hierarchical values, where each value contains one or more FTFs. ICECUP then applies this variable to every modal auxiliary in the corpus, classifying each one in turn using the FTFs, and storing the results.

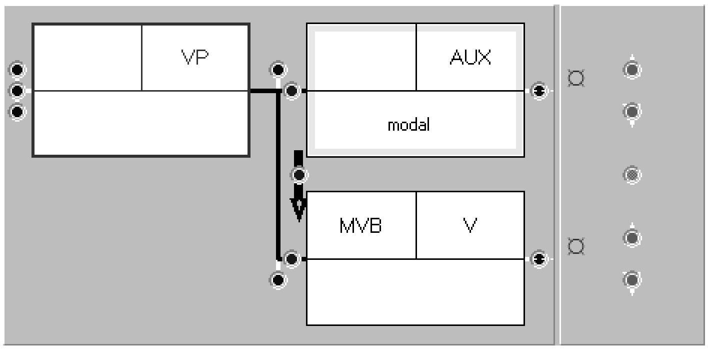

Figure 7
A1: An FTF for the pattern modal + main verb.

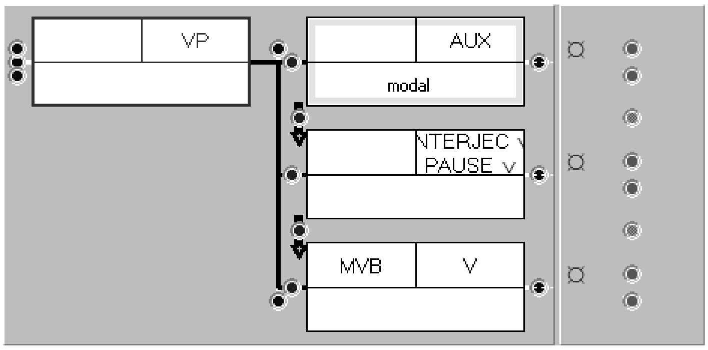

Figure 8
A2: An FTF for the pattern modal + X + main verb.
A3 is similar, but includes two 'X' nodes. A4 replaces the 'X' node with one or more semi-auxiliary verbs.

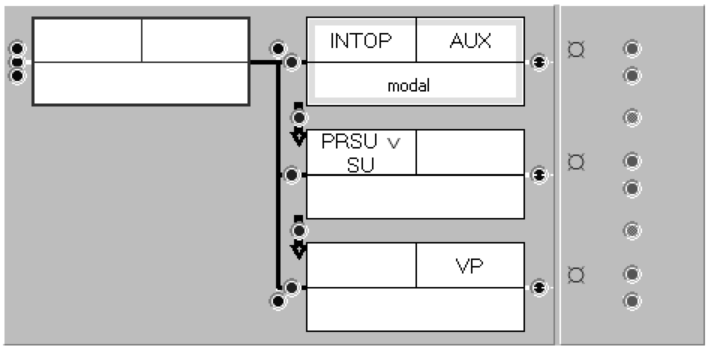

Figure 9
A5: An FTF for the pattern modal + subject + VP.
A6 and A7 include 'X' nodes before and after the subject.

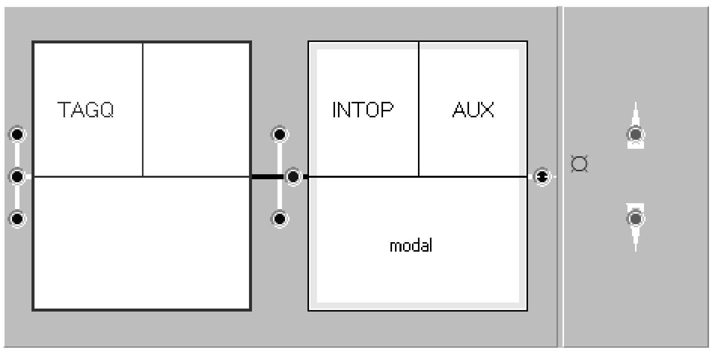

Figure 10
A9: An FTF for a 'tag question'.

6.1 Pattern A: modal + main verb

In DCPSE we searched for pattern A (modal + main verb) using eleven different sets of FTFs (A1–A11). In what follows each will be presented in turn.

A1 Declarative: modal + main verb

The simplest FTF for A1 is shown in Figure 7. This searches for all patterns in which a modal auxiliary verb is immediately followed by a main verb, as in the example below.

(1) That *should be* interesting. [DIA-10 #217]

The black arrow in the FTF signifies that the main verb node must follow the auxiliary verb immediately, i.e. without intervening constituents. Users can manipulate this arrow by clicking on the 'radio button' at its centre turning. One of the other options is to turn the arrow white, which indicates that the two nodes may be interrupted, for example by an adjunct. In this case we opted not to construct the FTF with a white arrow (coding for 'eventually follows') because this would also find e.g. *will have taught*, which would match pattern B.

We also include patterns where *not* follows the modal auxiliary verb. This is consistent with the treatment of modals with cliticised negative elements, e.g. *won't go*.[10] However, other negative elements qualify as intervening, e.g. the adverb phrase *never*, and these are considered as belonging to the next pattern.

A2 Declarative: modal + X + main verb

The FTF in Figure 8 represents a pattern in which a modal auxiliary is followed by an intervening constituent, which we will label 'X', and a main verb. For example:

(2) I'*ll* then *start* again. [DI-A01 #23]

ICECUP interprets black arrows as skipping over pauses, interjections, etc. Thus, for example, *I will* <pause> *do it* will match pattern A1.

A3 Declarative: modal + X + X + main verb

We extend the FTF in Figure 8 to include two 'X' nodes between the auxiliary and modal verb, and find a number of examples like the following.

(3) I'*ll* probably just *spend* Monday at the flat. [DI-C02 #130]

A4 Declarative: modal with semi-auxiliaries

Finally, we treat modal patterns which include semi-auxiliaries as a distinct class. In the TOSCA/ICE analysis the label 'semi-auxiliary' is more broadly applied than in Quirk *et al.* (1985), including a range of items described there (pp. 136–48) as intermediate between auxiliaries and main verbs (e.g. *ought to*, *had better*, BE *going to*, HAVE *to*, SEEM *to*). The TOSCA/ICE notation distinguishes between semi-auxiliaries followed by an infinitival verb form and 'semip' auxiliaries followed by an -*ing* form (e.g. KEEP in *I keep tidying it up*). This set covers both cases.

(4) So that I *wouldn't need to see* [DI-A12 #25]
(5) Then she said oh you *must stop doing* that [DI-A13 #150]

A5 Interrogative: modal + subject + VP

So far we have only considered declarative VPs containing modals and main verbs. FTFs A5–A8 match clauses in an interrogative pattern containing a preceding 'interrogative operator' (labelled 'INTOP'). An example is

(6) *Shall I go* first? [DI-A02 #2]
(7) *Will it matter* to you if he knows this is it? [DI-A11 #302]

We also include examples with 'provisional subjects' occupying the subject slot, such as *it* in (7) above which anticipates the underlined *if*-clause. Figure 9 reveals that the middle node matches either a subject ('SU') or provisional subject ('PRSU').

As in previous FTFs, the arrows in each case are black (i.e. the nodes follow each other sequentially without interruption) which means that nothing is permitted to intervene between the auxiliary and the subject, and between the subject and the verb phrase.[11] These restrictions ensure that examples like the following are not counted because they belong to a different pattern.

(8) Will you have been checking?

FTF A5 also finds an open interrogative pattern such as the following.

(9) What *would you take?* [DI-B12 #127]

A6 and A7 Interrogative: modal + X + subject + VP/Interrogative: modal + subject + X + VP

The FTFs for patterns A6 and A7 are similar to Figure 9 but we add an 'X' node between nodes. A6 and A7 differ only with respect to the position of 'X'. A6 returns no cases but A7 does obtain some examples:

(10) *Can you just sig*n that? [DI-B59 #171]
(11) How *can I best help* you over your careers advice? [DIA-08 #1]

A8 Interrogative: modal with semi-auxiliaries

There are a small number of cases of interrogatives with semi-auxiliary verbs.

(12) and then you know *will I have to unscramble* everything? [DI-J04 #132]
(13) *would you begin to relate*…? [DL-B22 #241]

A9 Tag questions

The FTF in Figure 10 identifies modals used in tag questions. These are reduced interrogative clauses which attach to a preceding clause, as in the following example:

(14) I suppose it would be optical scanning, *would it*? [DI-A05 #41]

A10 The 'code' pattern

So-called 'code' patterns are those in which an auxiliary verb is 'stranded', as in (15):

(15) *Yeah, I will.* [DI-A02 #137]

We use a simple FTF which matches cases where the auxiliary verb is in VP-final position.

A11 Non-interrogative inversion

Finally we have cases of inversion in declarative constructions, such as *So would I* (DI-B03 #264). The auxiliary is given the function *inverted operator* ('INVOP') in DCPSE, and cases are easily identified.

The figure below displays the distribution of the various subtypes of pattern A per million words in DCPSE. Over 86 percent of modal auxiliaries are to be found in patterns A1 to A11, and nearly 86 percent of these (10,656, excluding inversion and 'code') are in declarative patterns. If we examine declarative patterns A1 to A3, we find that the probability of adding one non-auxiliary element 'X' to a declarative pattern is around 12.4 percent (1,276 out of 10,281) but the probability of adding a second falls to 5.4 percent (69 out of 1,276). This fall in probability is significant: it implies that once one element is added it becomes more difficult to add a second.

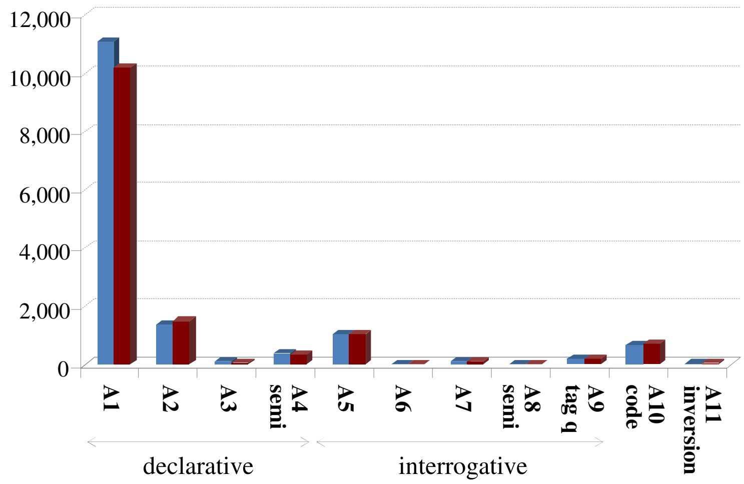

Figure 11
Distribution of pattern A in DCPSE (pmw), i.e. excluding perfect, passive and progressive auxiliary verbs.

Semi-auxiliary patterns are quite rare, but they seem to be *particularly* unusual in interrogative modal patterns. Only 5 out of 1,146 (0.43 percent) of the modal interrogative patterns contain a semi-auxiliary, compared to 2.87 percent of the declarative patterns. (This is a significant difference.) Finally, A10 ('code') corresponds to a little less than 5 percent of pattern A.

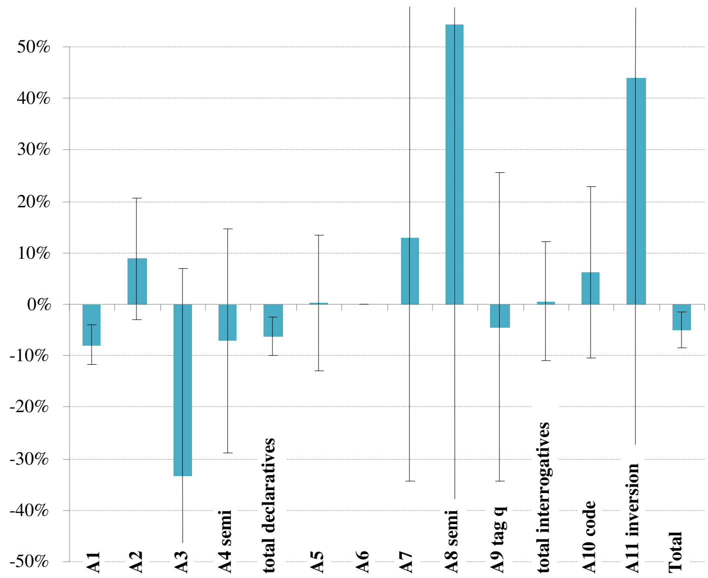

Figure 12
Absolute change (pmw) within modal pattern A in DCPSE, expressed as a percentage of 1961 data.
Some confidence intervals (longer = lower confidence) are truncated for reasons of space.

Table 4 summarises the results, including raw and per-million-word frequency in the 1960s and 1990s data, and absolute and relative change over time. In the table we calculate $d^{\%}$ first as an absolute proportion of all words[12] (as per Figure 12) and second, based on changes in the relative proportion of each category to the total. This second calculation allows us to differentiate a rise or fall in usage of a particular pattern from the overall trend.

Our data shows that the declarative patterns fall in the data set in absolute (per-million-word) terms by over 6%. This change appears to be driven by a decline in the simple declarative form A1, which falls significantly in both absolute and relative terms.

This fall is in marked contrast to the use of modals in interrogative patterns, tag questions and 'code' patterns, where changes are not significant. The decline of A1 and the relative increase in the declarative A2 indicates that declarative constructions of this pattern are tending to increase in length over time.

Table 4
Examining changes in the proportion of usage of patterns A1–A11 (modal + main verb).

	LLC (1960s)		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	$d^{\%}$	$d^{\%}$
Declarative						
A1 M + main verb	4,941	11,094	4,133	10,224	**-7.8%**	**-3.0%**
A2 M + X + main verb	607	1,363	600	1,484	+8.9% ns	**+14.6%**
A3 M + X + X + main verb	43	97	26	64	-33.4% ns	-29.9% ns
A4 declarative semi	166	373	140	346	-7.1% ns	-2.2% ns
Subtotal	5,757	12,926	4,899	12,119	**-6.2%**	-1.3% ns
Interrogative						
A5 M + SU + VP	468	1,051	426	1,054	+0.3% ns	+5.5% ns
A6 M + X + SU + VP	0	0	0	0		
A7 M + SU + X + VP	39	88	40	99	+13.0% ns	+18.9% ns
A8 interrogative semi	2	4	3	7	+65.3% ns	+73.9% ns
A9 tag question	90	202	78	193	-4.5% ns	+0.5% ns
Subtotal	599	1,345	547	1,353	+0.6% ns	+5.9% ns
Other						
A10 'code'	302	678	291	720	+6.2% ns	+11.1% ns
A11 'inversion'	13	29	17	42	+44.1% ns	+50.8% ns
Total	6,671	14,978	5,754	14,234	**-5.0%**	

We discuss these results in section 7. We now turn to pattern B.

6.2 Pattern B: modal + perfect HAVE + main verb

This pattern involves a modal verb and the perfect auxiliary *have.* We recognise ten sub-patterns, B1–B10. For reasons of space we will only show two FTFs in Figures 13 and 14; the FTFs for the other patterns look similar to those in section 6.1, with the relevant details changed.

B1 Declarative: modal + perfect HAVE + main verb

For pattern B1 the FTF in Figure 13 was used. Examples of this pattern are shown in (16) and, with the option of *not*, (17).

(16) It *might have been* a good degree. [DI-A06 #15]
(17) I *could not have worked out* the timetable. [DL-A04 #246]

We obtain the declarative patterns B2–B4 by adding ‘X’ nodes to Figure 13.

B2 Declarative: modal + X + perfect HAVE + main verb

(18) You see it *could so easily have been* a chapel. [DL-A07 #505]

B3 Declarative: modal + X + X + perfect HAVE + main verb

(19) At that time he *might also by chance have met* his sister Isabella. [DI-J14 #44]

B4 Declarative: modal + perfect HAVE + X + main verb

(20) Things *would have probably been* the same. [DI-A15 #146]

The corpus also contains one example of the pattern modal + perfect HAVE + X + X + main verb:

(21) I *should have like just whipped* up this amazing meal. [DI-B57 #314]

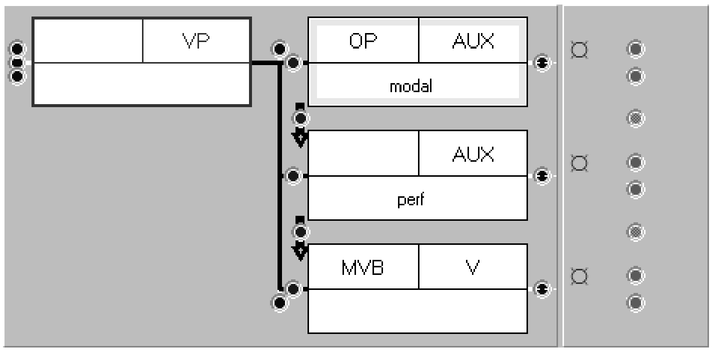

Figure 13
B1: An FTF for modal + perfect + main verb patterns.
B2 to B4 are similar, with added ‘X’ nodes.

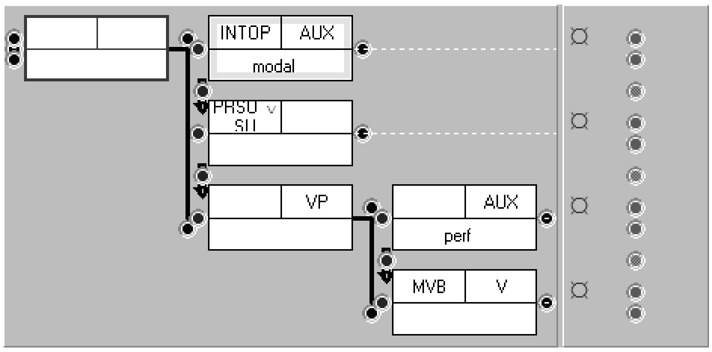

Figure 14
B6: An FTF for modal + subject + HAVE + main verb.
B7 and B8 include ‘X’ nodes at different locations.

B5 Declarative: modal with HAVE and semi-auxiliaries

There are a small number of examples of declarative modal patterns including perfect HAVE where semi-auxiliaries are also found. In all bar one case with *will*, the modal is *would*:

(22) they*’d have had to get* old too [DL-D02 #109]
(23) …my brother *would never have started drinking* [DL-D08 #111]

Interrogative patterns follow the framework outlined in Figure 14. Note that the perfect auxiliary precedes the main verb under the VP.

B6 Interrogative: modal + subject + perfect HAVE + main verb

There are numerous examples of the type shown below.

(24) How *would they have described* you? [DI-E06 #104]

B7 Interrogative: modal + X + subject + perfect HAVE + main verb

There are no examples of this pattern in the corpus, nor of cases where ‘X’ follows the perfect auxiliary.

B8 Interrogative: modal + subject + X + perfect HAVE + main verb

(25) But *could Britain really have acted* militarily in the summer? [DL-E04 #146]

B9 and B10 Interrogative: modal + subject + perfect HAVE + X + main verb/Interrogative: modal with HAVE and semi-auxiliaries

There are no examples of these patterns in the corpus.

Table 5 summarises the results. The number of cases of modal + perfect HAVE is falling in absolute (per million word) terms. This fall is replicated if we compare our data with a baseline of cases of perfect HAVE. In other words, out of all verb phrases containing the perfect auxiliary, the number of cases *where HAVE is preceded by a modal* falls over the same time period.

Within the set of patterns, only B1 stands out. It accounts for the lion's share (88%) of the cases, and the real-terms fall is significant. It cannot be distinguished from the overall trend further, that is, all we can say is that the results are explicable by the observed decline in B1 (and, thus, the set of declarative patterns as well).

Patterns B1 to B3 show the same type of significant fall in the probability of adding 'X' nodes as patterns A1 to A3 (the probability of adding an 'X' node declines from 7.4 percent to 1.7 percent).

Table 5
Examining changes in the proportion of usage of patterns B1–B10 (modal + perfect HAVE + main verb).

Perfect HAVE	**LLC (1960s)**		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	***d%***	***d%***
Declarative						
B1 M + perf + main verb	450	1,010	292	722	**-28.5%**	-2.8% ns
B2 M + X + perf + main verb	33	74	25	62	-16.5% ns	13.5% ns
B3 M + X + X + perf + main verb	0	0	1	2		
B4 M + perf + X + main verb	9	20	9	22	10.2% ns	49.9% ns
B5 declarative semi	5	11	4	10	-11.9% ns	19.9% ns
Subtotal	497	1,116	331	819	**-26.6%**	-0.2% ns
Interrogative						
B6 M + SU + perf + VP	10	22	8	20	-11.9% ns	19.9% ns
B7 M + X + SU + perf + VP	0	0	0	0		
B8 M + SU + X + perf + VP	1	2	0	0		
B9 – B10	0	0	0	0		
Subtotal	11	25	8	20	-19.9% ns	9.0% ns
Total	508	1,141	339	839	**-26.5%**	

6.3 Pattern C: modal + progressive BE + main verb

The patterns for C and D are variations on those for B. We replace features in FTFs specifying the perfect auxiliary node in B (see Figures 13 and 14), with those for progressive and passive respectively. We find examples like the following:

C1–5 Declarative: modal + progressive BE + main verb

(26) The LSE *would be doing* that principally. [DI-A05 #87] — **C1** M + BE + main verb
(27) We'*ll just be looking* at one small aspect. [DI-B87 #14] — **C2** M + X + BE + main verb
(28) He *might as well it seems be hitting* the wall outside. [DL-F03 #385] — **C3** M + X + X + BE + main verb
(29) …they *will be proudly taking* their places. [DL-F06 #279] — **C4** M + BE + X + main verb
(30) …it *could be beginning to backfire*. [DI-D19 #45] — **C5** M, BE, semi

C6 Interrogative: modal + subject + progressive BE + VP

(31) But *won't people be quibbling* about which band they're in? [DI-D14 #17] — **C6** M + SU + BE + VP

There are no examples of patterns C7 – C10 (i.e. interrogatives with additional nodes) in the corpus.

The results do not show a significant change over time, whether the trend is measured against the number of words in the corpus or as a proportion of progressive VPs. As with previous cases, a very high proportion of cases (86%) fall into the simple modal + progressive BE + main verb pattern.

Table 6
Examining changes in the proportion of usage of patterns C1–C10 (modal + progressive BE + main verb).

Progressive BE	**LLC (1960s)**		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	***d*%**	***d*%**
Declarative						
C1 M + prog + main verb	105	236	99	245	+3.9% ns	+0.1% ns
C2 M + X + prog + main verb	8	18	9	22	+23.9% ns	+19.4% ns
C3 M + X + X + prog + main verb	1	2	0	0		
C4 M + prog + X + main verb	2	4	1	2	-44.9% ns	-46.9% ns
C5 declarative semi	0	0	3	7		
Subtotal	116	260	112	277	+6.4% ns	+2.5% ns
Interrogative						
C6 M + SU + prog + VP	5	11	2	5	-55.9% ns	-57.5% ns
C7 – C10	0	0	0	0		
Subtotal	5	11	2	5	-55.9% ns	-57.5% ns
Total	121	272	114	282	+3.8% ns	

6.4 Pattern D: modal + passive BE + main verb

Here are some example sentences for pattern D.

D1–5 Declarative: modal + passive BE + main verb

(32) Uhm, they *may be allowed* into dance therapy. [DI-A01 #112] — **D1** M + BE + main verb
(33) They*'ll* <u>*probably*</u> *be done* in that evening. [DI-C09 #169] — **D2** M + X + BE + main verb
(34) …energies that I think *should* <u>*perhaps*</u> <u>*more fruitfully*</u> *be used* in other directions. [DI-D16 #24] — **D3** M + X + X + BE + main verb
(35) …I think emotion *should be* <u>*anyhow*</u> *shaped* by thought. [DL-A01 #143] — **D4** M + BE + X + main verb
(36) And the race *may* <u>*have to*</u> *be stopped* here. [DI-F10 #142] — **D5** M, BE, semi

D6–9 Interrogative: modal + subject + passive BE + VP

(37) … *should it be allowed* on a poster? [DI-B55 #143] — **D6** M + SU + BE + VP
(38) *Should that,* <u>*as Lord Scarman has suggested*</u>, *be changed*? [DI-D13 #50] — **D8** M + SU + X + BE + VP
(39) *Will it be* <u>*inevitably*</u> *postponed?* [DI-101 #122] — **D9** M + SU + BE + X + VP

The results are very similar to those for the progressive. We do not find a significant change over time in the use of modal + passive patterns, whether measured as a proportion of words or as a proportion of all passive uses, and variation within the set is not significant. Again, 81% of cases are found in the simple declarative pattern.

Table 7
Examining changes in the proportion of usage of patterns D1–D10 (modal + passive BE + main verb).

Passive BE	**LLC (1960s)**		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	***d%***	***d%***
Declarative						
D1 M + pass + main verb	249	559	209	517	-7.5% ns	-4.6% ns
D2 M + X + pass + main verb	22	49	28	69	+40.2% ns	+44.6% ns
D3 M + X + X + pass + main verb	1	2	2	5	+120.3% ns	+127.3% ns
D4 M + pass + X + main verb	10	22	11	27	+21.2% ns	25.0% ns
D5 declarative semi	7	16	5	12	-21.3% ns	-18.8% ns
Subtotal	289	649	255	631	-2.8% ns	0.3% ns
Interrogative						
D6 M + SU + pass + VP	9	20	7	17	-14.3% ns	-11.6% ns
D7 M + X + SU + pass + VP	1	2	0	0		
D8 M + SU + X + pass + VP	0	0	1	2		
D9 M + SU + pass + X + VP	1	2	1	2	+10.2% ns	+13.6% ns
D10 interrogative semi	0	0	0	0		
Subtotal	11	25	9	22	-9.9% ns	-7.0% ns
Total	300	674	264	653	-3.0% ns	

6.5 Patterns E–H

Patterns E–H yield very few data from the corpus. We present the results of our searches in the table below. In each case we searched only for the main patterns, without intervening units. We include interrogative and declarative patterns in our FTFs.

Table 8
Examining changes in the proportion of usage of patterns E–H (modal + multiple auxiliary verbs + main verb).

Multiple auxiliaries	**LLC (1960s)**		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	***d%***	***d%***
E M + prog + pass + main verb	1	2	0	0		
F M + perf + prog + main verb	10	22	1	2	**-89.0%**	-78.4% ns
G M + perf + pass + main verb	30	65	18	45	-31.6% ns	30.7% ns
H M + perf + prog + pass + main verb	0	0	0	0		
Total	41	92	19	47	**-47.7%**	

The table shows a per-million-word fall in the number of multi-auxiliary modal patterns in our data over the period from the 1960s to 1990s. (The same pattern is found if we compare against a VP baseline.) Note the concentration of examples which include the perfect auxiliary HAVE. This result appears consistent with the observation that perfect HAVE in modal combinations is in decline (see Table 5).

7 Discussion

7.1 Diachronic analysis of modal patterns

In this section we discuss the results of our research into diachronic changes in the use of modal patterns.

Turning first to Table 4, the results show that of the 14,316 modals in the DCPSE corpus the vast majority (12,425 or 86 percent), occur in pattern A, which involves a modal auxiliary followed by a main verb, with or without intervening lexical material. A grand total of 1,697 patterns, or just under 12 percent, involve a modal verb and at least one other auxiliary verb (e.g. *will have arrived, will have been examined*). Some 144 cases could not be classified, usually because they were marked as 'incomplete'.

What this table also reveals is that the overall decline of core modal usage by 6.4 percent (normalised per million words), as shown in Table 2, is potentially misleading. The change in usage over time varies according to which core modal we are discussing, and according to which patterns these modals occur in: different patterns decline at varying rates. In other words, changes in modal usage appear to correlate with clause structure. More specifically, we see that only in patterns A and B can significant changes be detected.

In Table 4 we see that Pattern A is falling in real terms by 5.0 percent. The sub-pattern A1 declines at a faster rate than Pattern A as a group (i.e. its *relative* decline is also significant). We can thus say that the simple form A1 'leads the way' or 'explains' the overall decline in Pattern A. Had A1 not declined, pattern A would not have fallen in frequency. The slightly longer A2 sub-pattern increases its 'share', so declarative patterns appear to be becoming longer. In Table 5, pattern B1 falls in real terms, and this appears to be the main factor causing the overall number of B patterns to decline.

What these cases have in common is that they are all declarative patterns. As we have seen, declarative cases account for between 80 and 90 percent of cases in each group. Therefore even when the percentage swing is small (7.8 percent in the case of A1) the decline may be statistically significant. There are no significant changes in usage within the interrogative structures. This may mean that no change is taking place or that we need more data in order to identify a trend. One way of addressing this within the dataset is to aggregate all interrogative modal patterns together. It turns out that the proportion of modals that are interrogative (compare 'INTOP,AUX(modal)' out of 'AUX(modal)', see Figure 2 for the FTF) does not vary significantly over time across DCPSE sub-corpora. This seems to indicate that the interrogative pattern has an extremely stable share of the modal set.

Finally, we commented that the perfect auxiliary HAVE is in decline in modal patterns. Table 9 shows that this is not simply due to a decline in the use of the perfect auxiliary in general. It is true that perfect HAVE is falling by 8.1 percent in absolute terms, but cases of HAVE preceded by a modal auxiliary are falling further. We can see that the proportion of cases of HAVE preceded by a modal fell by more than 20 percent over this time period (Relative column). This general decline appears to be reflected in Table 8 where we see the decline in use of longer patterns involving the perfect. Table 9 demonstrates that these longer cases are falling faster *still* (more than 43 percent, Relative column, top) than the 'modal + perfect + main verb' cases.

Table 9

Changes in the use of the perfect over time. From bottom to top: all perfects, perfect with a preceding modal auxiliary verb, modal + perfect followed by a second non-modal auxiliary verb (all $d^{\%}$ results are significant at $p<0.05$).[13] Note that this table shows relative change in proportion to the immediate superset in the row below (hence the double lines between rows).

Perfect	**LLC (1960s)**		**ICE-GB (1990s)**		**Absolute**	**Relative**
	raw	**pmw**	**raw**	**pmw**	$d^{\%}$	$d^{\%}$
M + perf + aux + main verb	40	90	19	47	**-47.7%**	**-43.1%**
M + perf + main verb	547	1,228	364	900	**-26.7%**	**-20.2%**
All perf	5,151	11,565	4,298	10,632	**-8.1%**	

7.2 Synchronic analysis of modal patterns

As well as examining change in frequency of patterns, we can also compare the distribution of sub-patterns from A to D to investigate whether and how they may differ synchronically, i.e. in terms of their total distribution. In reporting our data we have already made a number of observations about the synchronic distribution of each pattern. In this section we compare patterns within A–D with the overall distribution.

As usual, it is helpful to start with a more general categorisation, and then focus our attention more precisely in subsequent stages. A simple evaluation examines the declarative/interrogative distinction for each pattern A–D and compares this with the total. We find that the proportion of cases in the declarative form significantly increases when an additional perfect, progressive or passive auxiliary is added. Each of these cases is significantly higher in proportion than the total. If we carry out a pair-wise comparison with a 2×2 χ^2 test, we find that each (B, C, D) is distinguished from the simple form (A), but cannot be separated from each other.

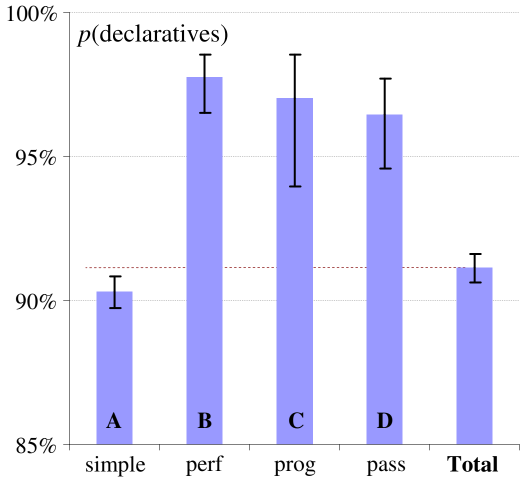

Figure 15

Proportion of declaratives, with 95% Wilson intervals (Wallis 2010), all significantly different from the overall pattern.

In order to break down the categories further we will need to concatenate some patterns under broad headings. Categories B2 and B4 are grouped into a single 'M + X + main verb' pattern (i.e. we do not distinguish where the other element, 'X', appears), ditto for C and D; and we similarly concatenate all 'M + X + X + main verb' patterns. The same approach is used to group interrogative patterns. Finally, as these are not easily included within the perfect, progressive and passive categories (or are likely to be subject to different pressures), we now exclude the 'short A' patterns (inverted declarative, 'code' and tag questions) from our analysis. Results are listed in Table 10.

Table 10

Contrasting the synchronic distribution of modal patterns (excluding 'code', tag questions etc.). Each distribution significantly differs from the total. Percentage differences from the overall distribution are listed where *particular* differences are significant (p<0.05).

	A: simple		**B: perf**		**C: prog**		**D: pass**		**Total**
M + main verb	**9,074**	-1%	**742**	+11%	**204**	+10%	458		10,478
M + X + main verb	1,207		76		20		71		1,374
M + X + X + main verb	69		1		1		3		74
declarative semi	306		**9**	-57%	3		12		330
M + SU + VP	**894**	+9%	**18**	-70%	**7**	-58%	**16**	-60%	935
M + SU + X + VP	79		1		0		4		84
interrogative semi	5		0		0		0		5
Total	11,634		847		235		564		13,280

A series of goodness of fit χ^2 tests, across all seven sub-patterns of A–D, reveals that the synchronic distributions for A–D are distinct from the overall (total) distribution.[14] To identify where precisely each distribution varies, we carry out 2-way goodness of fit comparisons for every cell (Wallis 2010) and cite the percentage difference from the expected value (the average value scaled proportionately). Thus we find that the simple perfect pattern B1 ('M + main verb') is 11 percent more frequent than the norm would lead us to expect.

This table can also be plotted as a proportion of the total of each pattern A–D, as in Figure 16 below. Our analysis indicates that detectable variation is concentrated in the two simplest categories (simple declarative, simple interrogative). In the main the other categories do not vary significantly.

Possibly the simplest explanation for our results is that the choice of interrogative or declarative construction has an impact on the speaker's decision to use an additional auxiliary. (We have already noted that the use of semi-auxiliaries also correlates with this choice.) In order to investigate further we have to focus on variation within the set of declarative patterns and within the interrogative patterns.

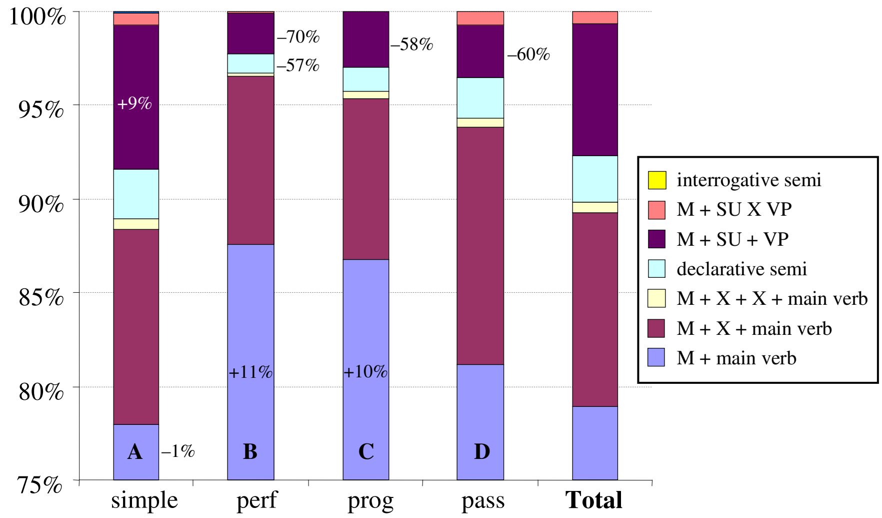

Figure 16

Synchronic distributions of modal patterns, expressed as percentages of the total, listing percentage difference from overall pattern ('Total' column) where significant ($p<0.05$). In each case over 75% are of the simple declarative 'modal + main verb' pattern.

The set of interrogative patterns are not subdivided further, as cases are almost exclusively of the simplest form. However, we can reanalyse Table 10 to focus exclusively on declarative constructions, eliminating the lower (interrogative) rows. This reveals that only the modal perfect HAVE distribution differs from the overall distribution. The declarative modal perfect shows a tendency towards simplification, i.e. the shorter declarative form is preferred (a relative increase of 5 percent) when the perfect auxiliary is included in a modal construction. Modal perfect patterns containing a further semi-auxiliary are also fewer (around 60 percent less) than expected.

8 Conclusion

Our research has a number of consequences for the way current change is researched. First, we use a corpus of *spoken* English, the *Diachronic Corpus of Present-day English* (DCPSE). This is particularly important given that spoken language is believed to be primary, that is, changes in the spoken language filter into the written language. Our evidence of spoken change could be seen as forecasting changes in written British English.

Secondly, measuring changes in terms of per-million-word frequencies generally gives a good indication of absolute change, but does not offer a complete picture. Ideally, one would wish to compare change in relation to alternates, or failing this, against a baseline for the choice under consideration (see Bowie, Wallis and Aarts forthcoming a). In this paper we have carried out a number of additional comparisons (against VPs for example) as a health-check on these absolute results. We have also examined relative change within the set. This is instructive because it allows us to differentiate between an overall trend and a 'trend within a trend' – as the perfect modal example above illustrates.

The experiments we have described in this paper would be inconceivable without a parsed corpus and a reliable well-defined query system like FTFs. We can gather data from the corpus very effectively, using FTFs at a range of specificity from the most general ('AUX(modal)') to the most detailed structured FTF.[15]

Our experiments have allowed us to differentiate between rates of change that particular modal patterns appear to be undergoing. Not all modal patterns appear to be undergoing change over this time period. The simple declarative pattern 'modal + main verb' (A1) appears to be falling the most dramatically over time, whereas most of the other patterns appear to be reasonably stable. This pattern is also synchronically the most common pattern. These facts are not necessarily independent.

The relative stability of the other patterns may be due to their possessing a particular marked semantic or idiomatic usage where alternate forms are not readily accessible to the speaker. A synchronic analysis of the same data shows that perfect, progressive and passive auxiliaries (and indeed semi-auxiliaries, see Section 6.1) tend to be more concentrated in declarative than in interrogative constructions, where simpler modal constructions predominate.

Within the B pattern (which involves perfect HAVE), the simple declarative B1 is also undergoing significant decline, but this cannot be differentiated from the fall in overall usage. However, by carefully contrasting the usage of modal perfect HAVE against non-modal usages of the perfect, we can see that the perfect is undergoing a decline at three levels (at least):

1) perfect HAVE is falling in absolute terms overall,
2) 'modal + perfect' cases are falling further, and
3) 'modal + perfect + other auxiliary' cases are falling further still.

We have also found that the declarative perfect appears to differ synchronically from the other patterns in tending to form shorter and simpler patterns than the average. On the perfect pattern, see Bowie and Aarts (forthcoming) and Bowie, Wallis and Aarts (forthcoming a). In, Bowie, Wallis and Aarts (forthcoming b) we further refine our research strategy by investigating changes in modal usage in different text genres.

References

Aarts, B. and S. A. Wallis. 2006. *The Diachronic Corpus of Present-Day Spoken English* (DCPSE). London: Survey of English Usage. [Software]

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

Aarts, Bas, Joanne Close and Sean Wallis. 2010. Recent changes in the use of the progressive construction in English. In B. Cappelle and N. Wada (eds.) *Distinctions in English grammar, offered to Renaat Declerck***.** 148-167. Tokyo: Kaitakusha.

Aarts, Bas, Joanne Close and Sean Wallis (forthcoming). Choices over time: methodological issues in current change. In B. Aarts, J. Close, G. Leech and S. Wallis (eds.) *The English verb phrase: corpus methodology and current change*. Cambridge: Cambridge University Press.

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

Bowie, Jill, Sean Wallis and Bas Aarts (forthcoming a). The perfect in spoken British English. In B. Aarts, J. Close, G. Leech and S. Wallis (eds.) *The English verb phrase: corpus methodology and current change*. Cambridge: Cambridge University Press.

Bowie, Jill, Sean Wallis and Bas Aarts (forthcoming b). Modals in genres.

Close, Joanne and Bas Aarts. 2010. Current change in the modal system of English: a case study of *must*, *have to* and *have got to*. In U. Lenker, J. Huber, and R. Mailhammer (eds), *The history of English verbal and nominal* constructions. Volume 1 of *English historical linguistics 2008: Selected papers from the fifteenth International Conference on English Historical Linguistics (ICEHL 15), Munich 24-30 August 2008.* Amsterdam: John Benjamins. 165-181.

Krug, Manfred. 1998. *Gotta* — The tenth central modal in English? Social, stylistic and regional variation in the British National Corpus as evidence of ongoing grammaticalization. In H. Lindquist, S. Klintborg, M. Levin and M. Estling (eds.) *The Major Varieties of English* (Papers from MAVEN 97). 177–191..Växjö: Acta Wexionensia.
Krug, Manfred G. 2000. *Emerging English modals: a corpus-based study of grammaticalization.* Berlin/New York: Mouton de Gruyter.
Leech, Geoffrey. 2003. Modality on the move: the English modal auxiliaries 1961–1992. In R. Facchinetti, M. Krug and F. Palmer (eds.) *Modality in contemporary English.* Berlin/New York: Mouton de Gruyter. 223-240.
Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith. 2009. *Change in contemporary English: a grammatical study*. Cambridge: Cambridge University Press.
Mair, Christian. 1997. Parallel corpora: a real-time approach to the study of language change in progress. In M. Ljung (1997)(ed.) *Corpus-based studies in English. Papers from the seventeenth International Conference on English Language research on computerized corpora* (ICAME 17), Stockholm, May 15-19 1996. 195-209. Amsterdam: Rodopi.
Millar, N. 2009. Modal verbs in TIME: frequency changes 1923-2006. *International Journal of Corpus Linguistics.* 14.2, 191-220.
Nelson, Gerald, Sean Wallis and Bas Aarts. 2002. *Exploring natural language: working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.
Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
Smith, Nicholas. 2003. Changes in modals and semi-modals of strong obligation an epistemic necessity in recent British English. In R. Facchinetti, M. Krug and F. Palmer (eds.) *Modality in contemporary English.* Berlin/New York: Mouton de Gruyter. 241-266.
Svartvik, J. (ed.) 1990. *The London-Lund corpus of spoken English: description and research.* Lund Studies in English 82. Lund University Press.
Wallis, Sean, and Gerald Nelson. 2000. Exploiting fuzzy tree fragments in the investigation of parsed corpora. *Literary and Linguistic Computing* 15.3, 339-361.
Wallis, Sean. 2008. *Getting Started with ICECUP IV. Experiments in Grammar and Lexis*. London: Survey of English Usage. www.ucl.ac.uk/english-usage/projects/next-gen/beta/g-start-4.pdf
Wallis, Sean. 2010. *z-squared: The origin and use of* χ^2. London: Survey of English Usage. www.ucl.ac.uk/english-usage/staff/sean/resources/z-squared.pdf

Notes

1. The results of this paper have been presented in Malta at the IAUPE 2010 conference, in Manchester at the symposium *Syntax: past and present* (2010), at the University of Lille 3 (2010), and in Logroño at the 2011 conference of the *Societas Linguistica Europaea*. We are grateful to the audiences for their comments. We acknowledge the support of the Arts and Humanities Research Council under grant AH/E006299/1.
2. For an overview of ICECUP 3.1 and IV see www.ucl.ac.uk/english-usage/resources/icecup. ICECUP can be downloaded with a sample corpus from DCPSE from www.ucl.ac.uk/english-usage/projects/dcpse/beta. The full DCPSE corpus on CD (plus software manual and detailed on-line help) can be obtained from our website.
3. Are researchers required to commit themselves to the Quirk phrase structure framework in order to use DCPSE? No. Linguists are entirely at liberty to disagree with the grammar used in the corpus. Note that we group categories (e.g. provisional subject and subject; semi and 'semip' auxiliary verbs) to perform our own analysis. The key question is *can linguists use the grammatical framework to reliably extract data?* ICECUP is designed around the premise that all results must be checked and double-checked against the corpus itself. ICECUP's interactive interface is designed around an exploratory mode of working with the data so that linguists can confirm results for themselves by examining matching sentences. Whereas we concentrate in this paper on reporting figures, a great deal of work has been undertaken in confirming that results are as sound and complete as possible.

4. Significance testing is carried out with a goodness-of-fit χ^2 test (Wallis 2010). This measures whether an observed change is significant compared with a given baseline, in this case the change in the number of words in the corpus.
5. Readers may be curious as to why changes in *shall* are statistically separable even though the percentage differences appear numerically very similar. The answer is that the **starting point** is different in each case. The difference separability test compares actual swings ($p_1 - p_2$) rather than *percentage* swings: in this case the two differences themselves differ sufficiently for the result not to be explicable by chance. The converse also applies – seemingly large differences in percentage swing may not in fact be statistically separable. For more on this and an alternative measure of change which is less problematic than $d^{\%}$, see Wallis (2010).
6. In DCPSE the modal auxiliaries are *can, may, shall, will, must* and their 'past tense' counterparts *could, might, should*, and *would*.
7. A confidence interval for $d^{\%}$ is a way of representing the degree of confidence we might have in our observed value, and can be plotted by so-called error bars or expressed as a numerical range (e.g. ±5%). We estimate that the actual value (in the population) is anywhere within the confidence interval at a given level of probability (say $p<0.05$). In our case the confidence intervals are computed using the optimal Newcombe-Wilson method (Wallis 2010).
8. Leech *et al.* (2009: 77, 284) found a decline in usage for the core modals of 11% in a corpus of spoken English of approximately 260,000 words, based on data from the Survey of English Usage. This percentage figure again excludes *ought* (*to*) and *need* (*to*).
9. We also cannot completely factor out variation in samples that might arise from differing modal-semantic orientation of particular texts (instructional texts expressing obligation, discursive texts expressing possibility, etc.) in each of the two subcorpora.
10. Results in which the adverb *not* intervened between the auxiliary and main verb were dealt with separately by employing a second FTF and reallocating the results. ICECUP IV makes this job much easier by serially classifying cases using multiple FTFs.
11. Apart from pauses and interjections, which ICECUP intelligently 'skips over'. See Nelson *et al*. (2002).
12. Comparing each trend against all VPs rather than words obtains a very similar distribution of results.
13. Numerical results are slightly different than those we have published elsewhere (Bowie and Aarts forthcoming, Bowie *et al.* forthcoming a) because in this experiment we did not exclude the stative idiom HAVE *got* (as in *he's got two kids*). However, the overall conclusions are broadly the same.
14. When we refer to 'expected' values here, we mean values in proportion to this overall distribution or compared to the average distribution (this amounts to the same thing).
15. This process of extraction is aided by ICECUP IV (Wallis 2008), a development of the standard exploration platform, which allows users to define classificatory variables containing as many distinct values as required, and each value can contain a logical combination of FTFs. These FTFs are applied in one pass, with the result that every single auxiliary in the corpus is classified by this variable. We then obtain frequency results from this 'combing' and sorting method.