Contemporary change in modal usage in spoken British English: mapping the impact of 'genre'

Jill Bowie, Sean Wallis and Bas Aarts
University College London

1 Introduction

In this paper we report on changing patterns of use of the core modal auxiliaries in present-day spoken British English. We examine variation over recent time (1960s–1990s), for the modals as a group and for individual modal auxiliary verbs.[1]

Previous studies have focused on written English and only occasionally have they considered the question of whether these changes are affected by text type. The default assumption appears to be that change over time in written English is representative of changes more generally and that, if text type makes a difference, the effect of this variation is minor. However, synchronic studies in both written and spoken English have shown that text type affects the likely use of a particular core modal. In this paper we therefore go one stage further: we investigate whether different spoken text types (sometimes referred to as spoken 'genres') yield different results.[2]

We use data from the *Diachronic Corpus of Present-Day Spoken English* (DCPSE), a corpus of British English material drawn from a range of spoken text types. Aarts, Wallis and Bowie (forthcoming 2012) report diachronic findings from DCPSE with a comparison of change in various modal verb phrase patterns (involving combinations with other auxiliaries, declarative and interrogative structures, and so on). This paper extends our investigations in another direction by examining whether *patterns of change over time* vary across different types of spoken text.

1.1 The modals in recent corpus studies

The English modal auxiliaries share a number of well-known grammatical properties that distinguish them from other auxiliary verbs. These include a lack of non-tensed forms, a lack of person–number agreement and occurrence with a following verb in bare infinitival form. The core modals surveyed in this paper are those modals which have been identified as 'central' on the basis of such grammatical properties: *can*, *could*, *may*, *might*, *shall*, *should*, *will*, *would* and *must* (e.g. Quirk *et al*. 1985: 137; cf. Collins 2009: chapter 2). Apart from *must*, these are usually paired as present and past tense counterparts of single lexemes (*can*/*could*, *may*/*might*, *shall*/*should*, *will*/*would*), although the relationships between the counterparts are complex. For present purposes it is useful to treat them as individual items, as each has its own profile of changing usage.

The aim of this paper is to survey changing usage in the grammatically delimited group of core modal auxiliaries. We therefore exclude several other modals identified by Quirk *et al*. as more 'marginal' members of the set in terms of their grammatical behaviour (*dare*, *need*, *ought to*, *used to*). We also exclude a range of 'quasi-modals', described by Collins (2009: 15) as 'periphrastic modal forms … formally distinguishable from, but semantically similar to, the modal auxiliaries', such as *have got to*, *have to*, *be to*, *had better*. See van der Auwera, Noël and Van linden (this volume) for a study of diachronic and regional variation in the use of *(had) better*, and for references to the literature on marginal modals and quasi-modals.

Recent diachronic trends in the use of the modals have been studied by Geoffrey Leech and colleagues, focusing mainly on written English. Their main findings are reviewed below; see Leech (this volume) for further discussion. Leech and colleagues also report some findings for spoken

[1] The research reported here stems from a project supported by the Arts and Humanities Research Council under grant AH/E006299/1. We gratefully acknowledge their support.

[2] In this paper (and despite the title) we prefer the term 'text type' as a neutral descriptor covering distinct sociolinguistic categories of text, since 'genre' is variously used and defined in the literature.

English, based on selected subsamples of DCPSE (whereas in this paper we report detailed findings drawn from the full corpus).

Leech (2003) and Leech *et al*. (2009) report findings from the 'Brown quartet' of corpora: four one-million-word corpora representing printed written British and American English from the 1960s and 1990s. The British corpora are the LOB (Lancaster–Oslo–Bergen) corpus (containing material from 1961) and its Freiburg counterpart FLOB (with 1991 material), while the American corpora are Brown (1961) and Frown (1992). These four corpora include a range of matching categories of printed text.

Leech *et al*. (2009) measure modal frequencies per million words (pmw), and report a significant overall decline of around 10% in modal frequency in both British and American English writing.[3] They also report results for individual modals. In British English significant declines are found for all core modals except *can*, *could* and *will*, which show no significant change. In American English the pattern is slightly different, with all core modals showing significant decline except *can*, *could* and *might*, which show no significant change.

Leech *et al*. (2009) also compare trends for the modals among four written 'subcorpora' groupings in the British English data. They find significant declines (ranging from about 8% to 14%) in three of these categories: press, general prose and fiction. In contrast, the learned category shows no decline (rather, a small non-significant rise). They do not give a breakdown of data by text type for individual modals. These four groupings in the corpora are more finely categorised into 15 text categories (for example, press comprises reportage, editorial and reviews). Leech (2003: 227) comments that, in the British English data, modal pmw frequencies decrease in 13 out of 15 categories, with the percentage decline ranging from 31.2% to 0.5%, and the remaining two categories are non-significant.

Leech (2011) has recently extended this diachronic study backwards to 1901 and forwards to the first decade of the 2000s, based on further (mainly written) corpora for both British and American English. The British English corpora (designed to match the earlier 'Brown quartet') are BLOB-1901 (covering 1898–1904), BLOB-1931 (1931), and BE06 (2005–8). The 1901 corpus is still incomplete, so Leech's initial results are based on a one-third representatively sampled cross-section. For American English, Leech reports results from COHA (*Corpus of Historical American English*), covering the 1910s to the 2000s, and COCA (*Corpus of Contemporary American English*), covering 1990–2010 and including spoken material (on these corpora, see Davies 2009, 2010). Leech's additional findings support the general picture of declining frequency of the modals from the middle of the twentieth century (British English) or earlier (American English), and show that significant decline continues into the 2000s in both varieties. He also finds, for both varieties, that the patterns of decline for a number of individual modals (such as *may*, *shall* and *must*) apply not just from 1961 to 1991 but over longer periods.

The importance of considering genre differences in diachronic change is highlighted when we compare these results with those of Millar (2009) in his restricted-genre study of modal verb frequencies in the *Time* magazine corpus. *Time* is, in his words, 'a popular American weekly news journal' (p. 197), and the corpus gives continuous coverage from 1923 to 2006, with around 1 to 1.5 million words per year. He reports results which in some ways differ surprisingly from those of Leech (2003). Using data grouped per decade, he finds a 22.9% *increase* in pmw frequencies of the core modals[4] from the 1920s to the 2000s — although the figures show a great deal of fluctuation over the decades. Comparing just 1961 and 1991 data to more closely resemble Leech's (2003) methodology, he finds a smaller increase of 2.81%.

[3] Leech et al. include in their study the more peripheral modals *ought (to)* and *need(n't)* in addition to the core modals listed above. However, as the frequencies of these forms are extremely low, they have little effect on the overall results (cf. the reanalysis of the data in Aarts, Wallis and Bowie forthcoming 2012). In terms of results for individual modals, these two are found to decrease significantly in British English, while their declines in American English are not significant (perhaps because their frequencies are already very low in 1961).

[4] Millar includes *ought* but not the very infrequent *need(n't)* in his study.

Millar does find, like Leech, clear patterns of decline for *shall* and *must*, both from 1961 to 1991, and over the longer term. However, he finds a strikingly different result for *may*. For this modal, Leech consistently finds patterns of decline. For American English, Leech finds a significant decline of around 32% in the Brown corpora from 1961 to 1991; a significant decline in COCA from the 1910s to the 2000s, with a mostly steady pattern of decline per decade; and a significant decline in COHA from 1990 to 2010, with a steady pattern of decline per five-year interval. In contrast, Millar finds an *increase* of 54% from 1961 to 1991, and his per-decade figures show a steady increase from the 1950s to the 1990s.

As Leech's data samples a wider range of genres, it seems reasonable to conclude, as does Leech (2011), that his findings are more indicative of patterns in the language as a whole, while Millar's contrasting finding for *may* appears to be specific to the genre of *Time* magazine. Millar himself notes an apparent 'surge in epistemic modality' in the magazine, which may reflect 'a shift towards greater speculation in reporting' (p. 213). The 'epistemic surge' is suggested by large increases in frequency not only of *may* but also of *could* and of modal adverbs such as *maybe*, and additionally by an increase in the *proportion* of occurrences of *may* which are epistemic.

Synchronic corpus studies have also exposed another factor which complicates the picture: the differing density of modals across text types. For example, Biber *et al*. (1999: 486–9) compare the registers of conversation, fiction, news and academic prose in the *Longman Spoken and Written English Corpus*, which includes both American and British English (they present graphs, but not exact figures or significance tests). They find that modals are most common in conversation and least common in news and academic prose. They report 'strikingly different distributions across registers' (p. 487) for most individual modals. For example, *can* and *will* are most common in conversation, while *may* is least common in conversation and most common in academic prose.

Finally, Collins (2009) presents a detailed synchronic corpus study of the modals and quasi-modals in British, Australian and US English, drawing data from the one-million-word British and Australian components of the *International Corpus of English* (ICE-GB and ICE-AUS), and from a smaller, specially assembled corpus of US English. He cites pmw frequencies for each lexical item, and for the ICE corpora provides a breakdown into four broad genre groupings: written printed, written non-printed, spoken dialogue and spoken monologue (while written and spoken figures are given for the US data).

We have carried out a simple re-analysis of Collins' data collating the nine core modal forms in ICE-GB (Appendix 1 presents raw figures). Contrasting spoken and written frequencies, we find that the overall modal pmw density is not significantly different,[5] but individual modals vary significantly. *Can*, *could* and *would* are more frequent in spoken English, whereas *may*, *must* and *should* are less frequent.[6] Within the written texts, the printed:non-printed distinction obtains a similar pattern, with no overall significant change, but *can*, *will* and *would* are more frequent in non-printed texts and *may*, *must* and *should* are more frequent in printed texts. This variation between individual modals is assessed as a proportion of the set of modals, i.e. by considering whether an individual modal changes its share of the total number of modals used.

The pattern of variation is rather different when we focus on Collins' spoken groupings. There are 26% (±3.5%) fewer modals per million words in the monologues than the dialogues. The share of the modal set taken up by individual modals is also different across spoken genres. *May* and *will* have a significantly larger share (110% ±30% and 22% ±8% respectively) of modals in the monologues than in the dialogues, whereas *would*'s share is significantly smaller (-30% ±6.5%).[7]

Of the four groupings, dialogues have the highest pmw frequency of modals while monologues have the lowest, and for modal use at least, spoken genre variation appears to be wider

[5] Using a 2×1 'goodness of fit' chi-square test at an error level $p<0.05$.

[6] This analysis is by share of set, although as the overall frequency is stable, the outcome would be similar were we to use a pmw baseline. We return to these concepts in our main exposition (see in particular section 3).

[7] These '±' figures are 95% (i.e. error level $p<0.05$) confidence intervals, i.e. we are 95% confident that the true population value is within this range. See Section 1.3.

than written. Given that corpus research has tended to focus on written data, due to its availability, this suggests that a focus on spoken English would be rewarding.

ICE-GB is a synchronic corpus collected over a brief timespan (1990–92). DCPSE, the corpus used in the present study, provides the opportunity to study change over time and how such change might differ between spoken text types. This corpus is described in the next section.

1.2 DCPSE

The *Diachronic Corpus of Present-Day Spoken English* (DCPSE) is a parsed (fully grammatically analysed) corpus of over 849,000 words of spoken British English. DCPSE is, to our knowledge, the largest available single corpus of parsed and checked spoken English. It comprises two subcorpora containing material from different periods. One subcorpus contains texts from the *London–Lund Corpus* (LLC) dating from the late 1950s to the 1970s (Svartvik 1990); the other, texts from the *British Component of the International Corpus of English* (ICE-GB) collected in the early 1990s (Nelson *et al.* 2002). The corpus therefore allows the investigation of recent change across a period of around thirty years.

The spoken data is divided into 'parsing units' (roughly, sentences), each of which is grammatically analysed in the form of a labelled phrase structure tree. An example is given in Figure 1, where the tree is displayed branching from left to right for compactness. Each node of the tree is labelled in terms of (i) function (top left, e.g. subject, noun phrase head), (ii) category (top right, e.g. noun phrase, adverb), and (iii) additional features where applicable (bottom section, e.g. singular, intransitive).

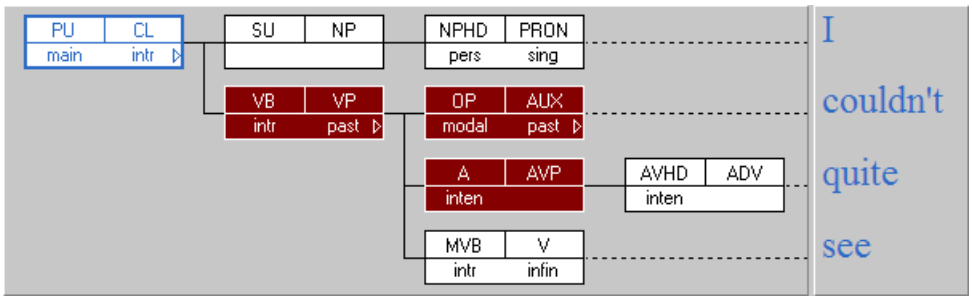

Figure 1. An example of a tree diagram in the DCPSE corpus, *I couldn't quite see* DI-F05 #6.[8]

PU = parsing unit, CL = clause, intr = intransitive, SU = subject, NP = noun phrase, NPHD = noun phrase head, PRON = pronoun, pers = personal, sing = singular, VB = verbal, VP = verb phrase, OP = operator, AUX = auxiliary, A = adverbial, AVP = adverb phrase, inten = intensifying, AVHD = adverb phrase head, ADV = adverb, MVB = main verb, V = verb, infin = infinitive.

This parse analysis is not merely descriptive. The entire treebank can be systematically searched and explored using software called ICECUP 3.1 (*International Corpus of English Corpus Utility Program 3.1*), which was specifically developed for the exploration of parsed corpora. This means that we can reliably identify instances of modals and – subject to the accuracy of the analysis – limit results to those within a particular grammatical context.

ICECUP has a flexible facility for searching grammatical structures. So-called *Fuzzy Tree Fragments* or 'FTFs' (Aarts *et al.* 1998; Nelson *et al.* 2002) are a diagrammatic query representation of search elements in a tree. The idea is fairly intuitive: put a label like 'VP' in a category slot and the query will obtain verb phrases, extend the query into a mini tree and that configuration must be found, and so on.

Figure 2 shows an example of an FTF used to search for any VP containing a modal auxiliary immediately followed by an adverb phrase. This will match examples such as the VP shown in Figure 1 (where the matching nodes are shaded).

[8] The text code DI-F indicates that this example is from the spontaneous commentary in the ICE-GB (1990s) subcorpus

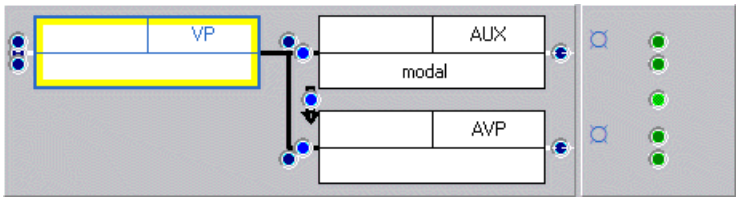

Figure 2. An example Fuzzy Tree Fragment. This pattern finds and matches VPs which contain a modal auxiliary immediately followed by an adverb phrase, such as the VP in Figure 1.

Figure 3 shows an example of a single-node FTF with lexical material specified. This FTF searches for any instance of the forms *would*, *'d*, or *wouldn't* classed as a modal auxiliary. The inclusion of grammatical category and feature information in searches for the modals excludes irrelevant instances of identical forms which would be retrieved in a simple lexical search (e.g. *'d* as a perfect auxiliary, *will* as a noun).

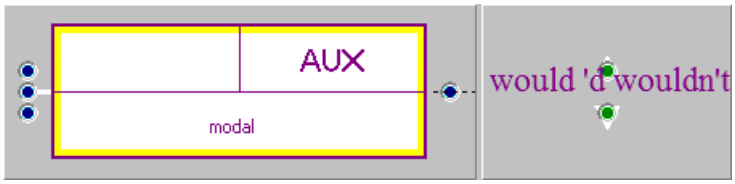

Figure 3. A FTF to find examples of modal auxiliaries taking one of the set of forms {*would*, *'d*, *wouldn't*}.

The spoken language data in DCPSE is sampled over a broad range of text types. The two subcorpora are sampled from ten matched categories. There are some unavoidable differences between the samples, however.

Note that LLC texts consist of 5,000 words each while ICE-GB texts are 2,000 words each, so that proportionately fewer texts were taken from the LLC than from ICE-GB to obtain similar subcorpus sizes. Table 1 shows the number of words and individual speakers per text category in the two subcorpora. The text categories also differ in size. Informal face-to-face conversation is the largest category, and the highly specialised category of legal cross-examination is the smallest.

text category	**LLC**		**ICE-GB**		**TOTAL**	
formal face-to-face conversations	46,291	(51)	39,201	(58)	85,492	(109)
informal face-to-face conversations	207,852	(146)	176,244	(398)	384,096	(544)
telephone conversations	25,645	(110)	19,455	(30)	45,100	(140)
broadcast discussions	43,620	(47)	42,002	(101)	85,622	(148)
broadcast interviews	20,359	(12)	21,385	(26)	41,744	(38)
spontaneous commentary	45,765	(50)	48,539	(60)	94,304	(110)
parliamentary language	10,081	(14)	10,226	(58)	20,307	(72)
legal cross-examination	5,089	(4)	4,249	(5)	9,338	(9)
assorted spontaneous	10,111	(8)	10,767	(5)	20,878	(13)
prepared speech	30,564	(14)	32,180	(71)	62,744	(85)
TOTAL	**445,377**	**(450)**	**404,248**	**(818)**	**849,625**	**(1,268)**

Table 1. Numbers of words and identified speakers (in brackets), in the two subcorpora of DCPSE.[9]

[9] The data presented in this paper have been obtained from a revised version prepared by the authors and others at the Survey of English Usage. The word counts exclude 'ignored' material, i.e. material excluded from the structural analysis because it represents nonfluencies such as repetitions and reformulations.

There are also differences in the time distribution of samples. The material in ICE-GB is from 1990–92, whereas LLC spans the years from 1958 to 1977, and the text categories in LLC are not evenly distributed across these years. For example, the single text in the legal cross-examination category comes from 1967, whereas texts within the parliamentary language and assorted spontaneous categories come from the 1970s. Other text types contain material from both decades in varying proportions. This means that while categories within LLC show some temporal variation as well as text type variation, each category can still be compared across LLC and ICE-GB as representative of earlier and later material *for that category*. We do, however, need to take care in attributing variation simply to 'time' or 'genre', as the two variables may not be strictly independent within the sample. Finally, the categories of spontaneous commentary, assorted spontaneous and prepared speech consist mainly of monologue.

1.3 Baselines for modal frequency

In the literature, as we saw in section 1.1, frequencies of modals are typically cited as per million words (pmw) as a method of 'normalisation'. However, a word frequency baseline is often far from optimal. When comparing results, the authors have argued (see e.g. Aarts, Close and Wallis, forthcoming 2012) that it is preferable to select a baseline for comparison which eliminates as much extraneous variation as possible. Why is this important?

Consider the question from the speaker's point of view. The decision to use a modal auxiliary verb represents a grammatical choice made within a verb phrase. Text categories may differ in their 'VP density', that is, in the frequency of VPs per million words. Ergo, the frequency of modals *per VP* is likely to obtain more reliable and precise results for changes in use than frequency per word. By a simple process of division we can factor out variation due to VP density.

Indeed, we can further narrow the baseline to tensed VPs (henceforth 'tVPs'), as modals cannot occur in non-tensed VPs (e.g. **He wants to can drive to work*, cf. *He wants to be able to drive to work*). Below we demonstrate that the text categories in DCPSE vary considerably in tVP density, and this underlines the importance of selecting an informative baseline.[10]

To search for tVPs in the corpus, the single-node FTF shown in Figure 4 was used. This FTF retrieves any auxiliary or main verb which bears the tense feature 'present' or 'past'. As each VP can include at most only one tensed AUX or V, this gives us a reliable count of tensed VPs in each category.[11]

Figure 4. A Fuzzy Tree Fragment which searches for auxiliaries or main verbs with the tense features 'present' or 'past'.

Table 2 lists pmw frequencies of tVPs in LLC and ICE-GB (the earlier and later subcorpora respectively), broken down by text category. (For raw frequencies see Appendix 2.) The table also shows the percentage change in frequency across the two subcorpora, and a 'goodness of fit' χ^2 test

[10] For frequencies of individual modals, the total set of modals provides another useful baseline. See Section 3.

[11] An alternative way to search for tensed VPs would be to search for VP nodes with the features 'present' or 'past'. Because of the parsing system used in the corpus, this method excludes certain instances where tensed AUX or V nodes occur directly under a clausal (CL) node rather than under a VP node, for example in tag questions (e.g. *He saw her, **didn't** he?*) and interrogatives where a copular verb acts as operator (e.g. ***Is** she angry?*). For present purposes, we prefer to use the more inclusive search, as the modal/non-modal distinction applies to these contexts (cf. *He could see her, **couldn't** he?*; ***Will** she **be** angry?*).

checks whether a change is a significant one. Figures 5 and 6 visualise pmw frequencies and the percentage differences in these frequencies, respectively.

Text category	LLC	ICE-GB	Change in frequency	
			%	χ^2
formal face-to-face	118,922	128,849	8.35%	**16.94**
informal face-to-face	134,312	136,589	1.70%	3.65
telephone conversations	125,795	130,558	3.79%	1.96
broadcast discussions	121,091	118,947	-1.77%	0.82
broadcast interviews	111,351	127,753	14.73%	**23.43**
spontaneous commentary	88,474	93,924	6.16%	**7.67**
parliamentary language	91,261	99,550	9.08%	3.66
legal cross-examination	145,412	114,851	-21.02%	**16.45**
assorted spontaneous	129,562	101,700	-21.50%	**35.14**
prepared speech	107,152	88,875	-17.06%	**53.55**
Total	122,348	122,232	-0.09%	0.02

Table 2. Tensed VPs pmw, by text category, compared across the two subcorpora of DCPSE. Values of χ^2 in bold indicate that the given percentage swing is significant (i.e. not zero) at $p<0.05$.

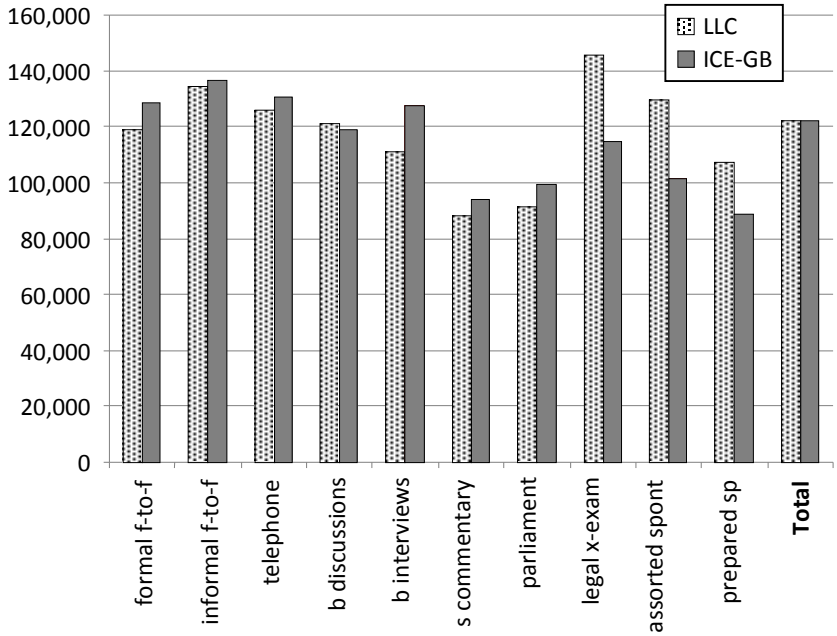

Figure 5. Synchronic and diachronic variation in evidence. The graph shows the frequency of tensed VPs pmw, by text category, across the two subcorpora of DCPSE.

In Figure 5 we simply plot the number of tensed VPs per million words in each text category and subcorpus (LLC and ICE-GB). Note that the labels for the text categories are abbreviated from the full forms given in Table 2 (e.g. 's commentary' is used for 'spontaneous commentary'). We can see that the 'Total' normalised frequency is almost perfectly constant over time, but the same cannot be said of individual text categories. It can also be seen that there is considerable variation in what we might call 'tensed VP density' (the rate of tVPs pmw) between text categories. For example, spontaneous commentary and parliamentary language have a relatively low tVP density in both LLC and ICE-GB.

In Figure 6 we further explore change over time by plotting percentage swing figures (the '%' column in Table 2) with 95% confidence intervals on these changes. This is a way of estimating

the accuracy of a given observation: a 95% interval means that there is a 1 in 20 chance that the observation would be outside this range were we to repeat the experiment. This is another way of expressing a significance test: if the interval does not cross the zero axis, the change is more or less than zero, which is another way of saying that the difference is 'significant'.

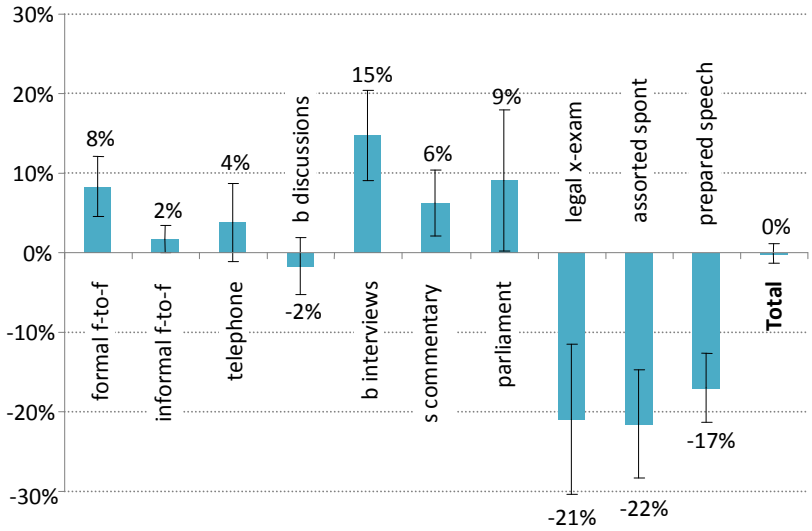

Figure 6. Percentage change in pmw frequencies of tVPs between LLC and ICE-GB subcorpora. The error bars are 95% confidence intervals. Where the line does not cross the zero axis, the change may be said to be significant (at $p<0.05$).[12]

Here again we see that, whereas tVP density overall ('Total' column) does not change significantly over time (i.e. from LLC to ICE-GB), the individual categories do show variation. Three categories exhibit a significant increase (formal face-to-face conversation, broadcast interviews, and spontaneous commentary) while another three show a significant decrease (legal cross-examination, assorted spontaneous, and prepared speech).

What is happening? There are two possible explanations, both potentially true, for this variation in tVP density over time. One is that this is the result of sampling differences; the other that it results from changes in language use within text categories. The particular reason does not concern us here, although it may be worthy of investigation in its own right.

This leads us inexorably to the conclusion that any comparison in terms of modal frequency set against a pmw baseline cannot address this source of variation. With a pmw baseline, we simply cannot say whether a change in the frequency of *modal tensed VPs* is due to a change in frequency of *tensed VPs*, or a change in the use of *modals*, or both. As tVP variation is significant and sizeable, a pmw baseline is likely to mislead.[13] We therefore factor out this extraneous variation by considering the frequency of modals per tVP instead of per million words.

1.4 Plotting change over time

At this point it is useful to briefly note that changes over time can be plotted by a number of different measures. Figure 6 plotted a change expressed in terms of percentage swing, which can be defined in the following way.

[12] We calculate accurate confidence intervals on these measures using Newcombe's (1998) method, giving us a more precise 2×2 χ^2 test of significance. Confidence intervals allow us to do two things: visualise certainty and test for significance at the same time. Throughout this paper we use a 95% confidence level (error level $p<0.05$) as standard.

[13] Recall that in section 1.1 we observed that Leech's and Millar's results, calculated per million words, showed a rise in modal use in the *Time* magazine corpus but a decline in the more broadly sampled Brown quartet. Since DCPSE subgenres show substantial variation in tensed VP density over time, it seems likely that similar variation would exist between *Time* and Brown, and we may hypothesise that the pmw baseline itself contributes to these apparently very different results.

First, we define **simple swing** as the difference:

$$swing = p_2 - p_1,$$

where p_1 and p_2 represent proportions from subcorpus 1, 2, etc. These proportions are proportions of a baseline, e.g. per million words, or per tensed VP. Ideally they should represent the probability of selecting an item (e.g. a modal tensed VP) from the set of items covered by the baseline (e.g. a tensed VP). **Percentage swing** is calculated as a fraction of the starting point, p_1:

$$percentage\ swing = (p_2 - p_1) / p_1.$$

The idea with percentage swing is that we obtain a single value which combines both an absolute starting point (p_1) and the change over time. While this is useful and fairly intuitive, and therefore commonly used (we quoted percentage swings in section 1.3, for instance), this measure has some important limitations.

Figure 6 presents positive or negative percentage swing from a zero axis: it does not show the start and end points of the change (cf. Figure 5) or the absolute size of the change. Percentage swing does scale the difference as a proportion of the starting point, which can be an advantage in comparing high and low frequency terms. However, it is difficult to compare the changes between two categories (such as broadcast interviews and spontaneous commentary) using percentage swing: if the starting point p_1 differs between these categories (as is likely), then the lengths of each bar (a function of p_1) are not on the same scale.

In Figure 7 we employ an alternative visualisation based on start and end points, p_1 and p_2, to plot the simple swing for the tVP data in Figure 5. The absolute height of columns in Figure 5 become start points and end points for any given change (e.g. legal cross-examination, indicated). The length of the bar represents the simple swing (difference), with the shading indicating the direction of the change, from LLC start points (lighter) to ICE-GB end points (darker).

Confidence intervals are calculated for the difference $p_2 - p_1$ using Newcombe's (1998) method, as previously. Thus in the case of legal cross-examination we can read this interval as saying that, at a 95% confidence level, the normalised frequency of tensed VPs falls by between 17,000 and 44,000 per million words. For convenience we attach the confidence interval to the end point (this also makes the direction of change clearer). This does not imply that the start point is certain and the end point uncertain, rather that *the difference between start and end is within this range*.

In order to decide which intervals represent a significant change, note that the relevant error bar is interior to the column, against the direction of change. Where the error bar does not cross the start point, the change is significant (at a 95% error level, i.e. $p<0.05$). Thus, in broadcast interviews, the frequency of tVPs increases significantly (from around 111,000 pmw in LLC to around 128,000 pmw in ICE-GB), whereas in prepared speech the frequency decreases significantly (from around 107,000 pmw in LLC to around 89,000 pmw in ICE-GB). Contrastingly, it can be seen, for example, that the increase in the telephone conversations category is not significant, as the error bar crosses the start point of the column.

This strategy allows us to capture a considerable amount of information in a single graph. Furthermore, it allows us to compare changes between categories which differ in their starting points. For example, we can clearly see that the values for broadcast interviews and spontaneous commentary are separated out, something that Figure 6 fails to illustrate.

In conclusion, both simple and percentage swing can be used to indicate the degree of observed change, and each metric has its advantages. Percentage swing has the advantage that it places differences of larger or smaller quantities on approximately the same scale, but it is not easy to compare terms which may be moving in opposite directions and it obscures differences in scale. For this reason, we use simple swing graphs in sections 2 and 3 below. In sections 4 and 5 we use percentage swing as figures become increasingly complex.

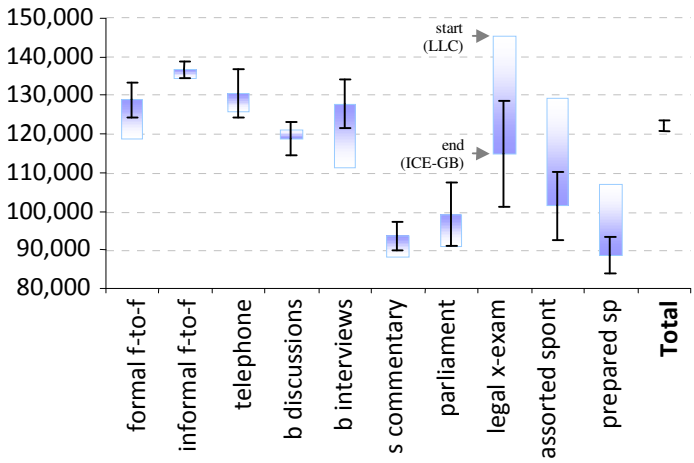

Figure 7. Plotting simple swing of tVPs pmw (from LLC to ICE-GB). Shading indicates the direction of travel, from LLC starting points (lighter) to ICE-GB end points (darker). Where the error bar does not cross the tail (starting point) of the column, the change is considered to be significant (at $p<0.05$).

2 All modals by text category over time

We report findings on the core modals *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will* and *would*. Negative and shortened forms (e.g. *won't* and *'ll*) are included in the data. To search for the modals we used a set of FTFs like the one shown in Figure 3 above, with lexical material specified.

First we will consider the frequencies of all core modals. We begin by comparing their pmw frequencies in the different text categories of the two subcorpora of DCPSE. We consider pmw frequencies to allow comparison (i) with other studies which report pmw frequencies; and (ii) with the results per tVP to be presented later. The pmw results are given in Table 3, and changes are visualised in Figure 8. The table gives both simple and percentage changes in frequencies, and the graph shows simple swing along with start and end points (cf. Figure 7). (For the raw frequencies of the modals, see Appendix 2.)

Text category	**LLC**	**ICE-GB**	**Change in frequency**		
			simple	%	χ^2
formal face-to-face	18,557	18,520	-37	-0.20%	0.00
informal face-to-face	17,464	18,049	584	3.35%	1.84
Telephone conversations	26,867	18,093	-8,774	-32.66%	**36.89**
Broadcast discussions	18,890	14,809	-4,082	-21.61%	**21.11**
Broadcast interviews	19,304	14,356	-4,948	-25.63%	**15.23**
spontaneous commentary	9,177	10,919	1,742	18.98%	**7.09**
parliamentary language	25,394	22,101	-3,294	-12.97%	2.32
legal cross-examination	11,004	22,358	11,354	103.18%	**18.46**
assorted spontaneous	18,198	11,702	-6,496	-35.69%	**14.82**
prepared speech	14,003	13,021	-983	-7.02%	1.12
Total	17,376	16,287	-1,089	-6.27%	**14.91**

Table 3. All modals pmw by text category, compared across subcorpora of DCPSE. The bold figures in the χ^2 column indicate significant changes over time (cf. Table 2).

Overall, the modals as a group show a significant fall of around 1,000 pmw or 6% of the LLC figure (see Total row, Table 3). However, this overall decline is not the end of the story.

We can see considerable variation between text categories. Spontaneous commentary has a comparatively low frequency of modal verbs, while parliamentary language shows a comparatively

high frequency. The effect of the diachronic dimension is also very different between categories. Significant decreases are found in four categories (telephone conversations, broadcast discussions, broadcast interviews, and assorted spontaneous speech), whereas significant *increases* are found in two categories (spontaneous commentary and legal cross-examination). These two text categories appear to 'buck the trend'. The obvious question is, what is going on?

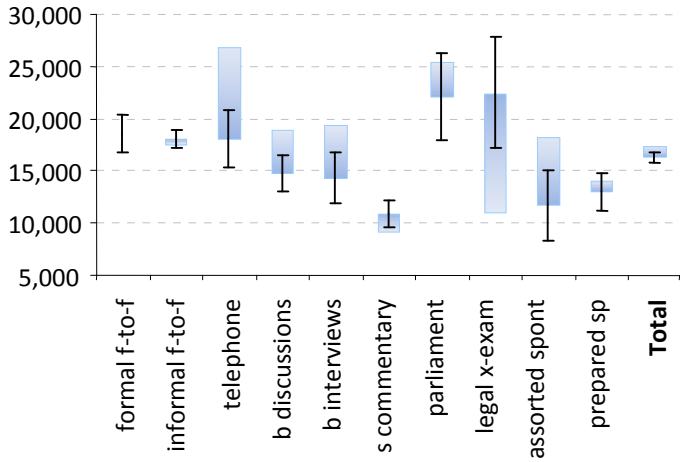

Figure 8. Modals pmw by text category, showing simple swing from LLC to ICE-GB.

We have already seen in section 1.3 that tensed VPs vary per million words by text category and time. As modal tensed VPs are a subcategory of tVPs, we first need to remove this variation from our analysis. We factor this variation out very simply by computing proportions p_1, p_2, etc. as a proportion of tensed VPs. This obtains the results shown in Table 4.

Text category	**LLC**	**ICE-GB**	**Change in frequency**		
			simple	%	χ^2
formal face-to-face	0.1560	0.1437	-0.0123	-7.89%	2.66
informal face-to-face	0.1300	0.1321	0.0021	1.62%	0.44
telephone conversations	0.2136	0.1386	-0.0750	-35.11%	**44.27**
broadcast discussions	0.1560	0.1245	-0.0315	-20.19%	**18.11**
broadcast interviews	0.1734	0.1124	-0.0610	-35.18%	**32.91**
spontaneous commentary	0.1037	0.1163	0.0125	12.07%	3.05
parliamentary language	0.2783	0.2220	-0.0563	-20.22%	**6.15**
legal cross-examination	0.0757	0.1947	0.1190	157.25%	**33.86**
assorted spontaneous	0.1405	0.1151	-0.0254	-18.08%	2.98
prepared speech	0.1307	0.1465	0.0158	12.10%	2.77
Total	0.1420	0.1332	-0.0088	-6.18%	**14.48**

Table 4. Modals per tVP by text category, compared across subcorpora of DCPSE. The χ^2 column compares the ratio of modals (LLC:ICE-GB) to the ratio of tVPs for significant difference.

This table shows the rates of modals per tVP. The overall rate is around 0.14 per tVP: that is, around 14% of tensed VPs contain a modal auxiliary verb. This rate for modals as a category is of interest in itself and more intuitively meaningful than a pmw rate.[14] The proportion of tVPs containing a modal auxiliary ranges widely across text type and time, from around 8% to 28%, as Figure 9 illustrates.

[14] Biber et al. (1999: 456) report a roughly similar rate of 10–15% in their four registers of conversation, fiction, news, and academic writing in the *LSWE* (*Longman Spoken and Written English*) *Corpus*, which includes both American and British English. (Their rate is for modals, which they consider as untensed forms, as a proportion of finite verb phrases; this in effect corresponds to our proportions.)

Table 4 reports simple and percentage changes in modal frequency per tVP from earlier to later subcorpora. The overall frequency declines significantly from LLC to ICE-GB by 0.0088 per tVP (simple swing). This is a *percentage* swing of -6.18% – an overall decline which is similar to that for pmw frequencies (-6.27%). (Recall that percentage swing is the difference between ICE-GB frequency and LLC frequency *as a proportion of LLC frequency*.) The figures cited in the text below are also percentage swing figures.

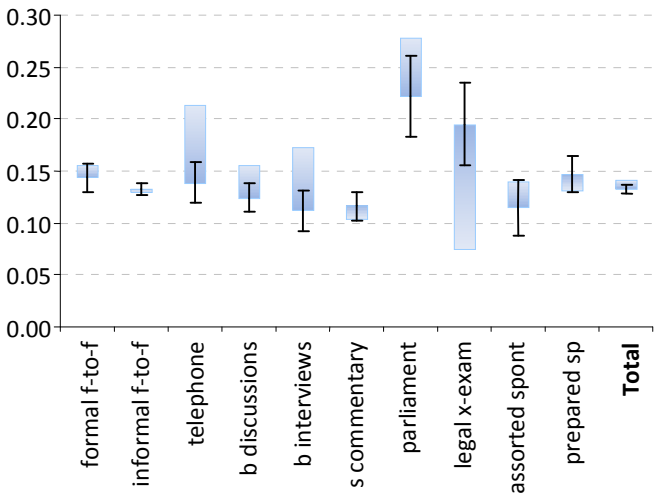

Figure 9. Modals per tVP by text category, showing simple swing from LLC to ICE-GB.

The proportion of tensed VPs containing a modal has declined significantly in four categories, and increased significantly in one (legal cross-examination). The 'assorted spontaneous' category no longer reports a significant change. For most categories, the results are similar to those obtained using pmw measures. Both methods find significant decreases in the categories of telephone conversations, broadcast discussions, and broadcast interviews, and a significant increase in the legal cross-examination category.

However, the *size of the change* observed differs considerably in some of these cases. For example, the decrease in the broadcast interview category is 35% (±11%) in the tVP results as against 26% (±13%) in the pmw results.[15] This category is one where tVP density pmw was found to increase significantly, hence the greater decline in modal frequency once this increase in tVP density is taken into account.

Elsewhere, the two baselines obtain different results. Spontaneous commentary shows a significant increase (19% ±14%) in the pmw results, but this increase is not significant in the tVP results; and as already noted, 'assorted spontaneous' shows a significant decrease (36% ±18%) in the pmw results, but the decrease is not significant in the per tVP results. The pmw results seem to be due to changing frequencies of tensed VPs pmw (an increase for spontaneous commentary, and a decrease for assorted spontaneous; see Figure 7), rather than to changing patterns of modal usage. On the other hand, parliamentary language exhibits a significant decrease (20% ±14%) in the tVP results, yet set against a pmw baseline, the decrease is not significant. It should be clear to the reader by now that this choice of baseline is extremely important, and, it seems, especially so when results are compared across text categories (and, by extension, between differently sampled corpora).

When considered per tensed VP, legal cross-examination is the only text category which shows a significant increase in modal frequency. The size of this increase is quite dramatic at 157% (±54%). However, this is the smallest category in terms of number of words (see Table 1) and is drawn from only nine speakers in total; the raw numbers of modals are also relatively small. Therefore these results should be treated with caution.

[15] These '±' figures represent approximate confidence intervals cited at the 95% level. See section 1.3.

We mentioned in the introduction that the text categories in the earlier subcorpus, LLC, are not evenly distributed across the 1960s and 1970s. It is worth noting, therefore, that the categories which show a significant decline are not all ones with a preponderance of earlier over later material: one (broadcast discussions) does have more earlier (1958–61) material than 1970s material, but the other three are weighted towards 1970s material and indeed one of them (parliamentary language) has only 1970s material.

It is also noteworthy that not all text categories show a decline in overall modal frequencies. No overall decline is seen in informal conversation, for example, much the largest text category in the corpus and one which might be expected to be at the head of changes unfolding in the language. To gain a fuller picture, it is necessary to look at the patterns of change among individual modals.

3 Individual modals over time

So far we have considered the frequencies of all core modals taken together. However, previous research has shown not only considerable synchronic variation in frequency among individual modals, but also varying diachronic trends over recent decades (e.g. Leech 2003; Leech *et al*. 2009; Aarts, Wallis and Bowie forthcoming 2012). We now turn to consider the frequencies of individual modals, first with all text categories taken together, and then (in the next section) within different text categories. Again, see Appendix 2 for modal frequency data.

Table 5 summarises the results for individual modals, against a tensed VP baseline. Simple and percentage swing per tVP are listed, along with the results of two 'goodness of fit' χ^2 tests. Column A evaluates whether each modal changes significantly against a tensed VP baseline. Column B considers each modal as a member of the overall set of modals, i.e. whether the modal significantly increases or decreases its relative proportion (share) of the overall set.[16]

Lemma	**LLC**	**ICE-GB**	simple	%	A: χ^2 (tVPs)	B: χ^2 (modals)
can	0.0297	0.0317	0.0020	6.87%	3.52	**13.53**
could	0.0157	0.0139	-0.0018	-11.18%	**5.36**	1.14
may	0.0072	0.0043	-0.0028	-39.77%	**36.18**	**27.51**
might	0.0055	0.0058	0.0003	5.17%	0.37	1.92
must	0.0086	0.0039	-0.0046	-54.19%	**87.85**	**73.50**
shall	0.0035	0.0018	-0.0017	-48.31%	**27.66**	**22.42**
should	0.0091	0.0078	-0.0012	-13.61%	**4.65**	1.48
will	0.0274	0.0309	0.0035	12.94%	**11.20**	**26.06**
would	0.0354	0.0329	-0.0025	-6.98%	**4.62**	0.06
Total	0.1420	0.1332	-0.0088	-6.18%	**14.48**	

Table 5. Proportions of individual modals per tensed VP, including negative and truncated forms. Columns A and B test for significant variation from tVPs and all modals respectively.

We can visualise these changes using simple swing graphs. Figure 10 plots simple swing against two baselines: as a proportion of tensed VPs[17] (using the left-hand axis) and as a fraction of the modal set (in effect, the set of modal tensed VPs, right axis). Thus we can see that the modal auxiliary *can* is found in around 0.03 of all tensed VPs and this figure does not increase

[16] Note that in employing a tVP baseline we do not claim that every tensed VP could take a given modal verb (*can*, *could*, etc.). We are not claiming that the baseline set represents a set of alternate forms, merely that it is a more reliable baseline than a word count. Similarly, in comparing *can*, etc., against the modal set we are identifying statistically significant variation within the modal set, rather than explaining this as arising by speaker choice.

[17] The results for percentage change measured per tVP are very similar to those reported in Aarts, Wallis and Bowie (forthcoming 2012: Table 2) for percentage change measured per million words (recall that, when all texts are taken together, tVP density does not differ significantly between the earlier and later subparts of DCPSE).

significantly, but its share of the modal set (i.e. of modal tensed VPs) does increase significantly over the same period.

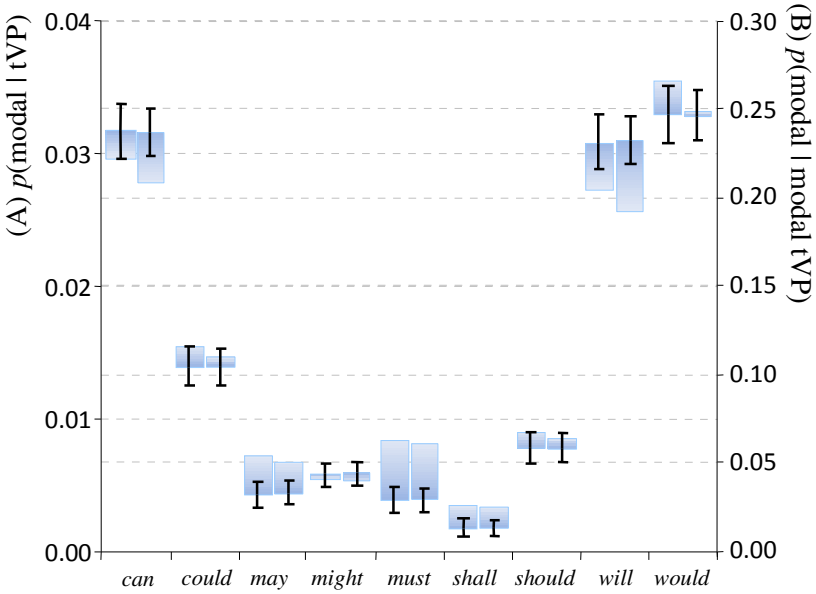

Figure 10. Changes in the use of the core modals (A, left axis) as an absolute proportion of tensed VPs and (B, right axis) as a relative proportion of the set of modals, measured as simple swing from LLC to ICE-GB.

Figure 10 and Table 5 also confirm the expected finding that frequencies of individual modals vary synchronically in this corpus, with *will*, *would*, and *can* being by far the most frequent. Each of those items comprises over 20% of the total modal figures, while *could* comprises over 10%.

Examining diachronic change over time, considered per tVP, *may*, *must*, and *shall* show significant and substantial declines in use (ranging from 40% to 54%). We might say that these modals are 'atypical' in that they show trends which are significantly different from the overall trend of decline for the modal set (cf. Table 5, Column B, and Figure 10). That is, not only are they declining in use in circumstances where they might be used, they are also decreasing their share of the modals.

By contrast, the modals *could*, *should*, and *would* all show significant declines of a smaller magnitude (from 7% to 14%), and they are 'typical' modals in the sense that these trends are not statistically separable from the behaviour of the modal set. (In other words, they do not significantly change as a proportion of the modal set.)

Only one modal, *will*, exhibits a significant increase in frequency per tensed VP (13% ±7.5%). The increase for *can* per tVP is not statistically significant. However, both *can* and *will* behave 'atypically' compared to the modal set and increase their share significantly (note the contrasting results for *can* in Columns A and B of Table 5, and see also Figure 10). Finally, *might* shows no significant change, whether measured per tensed VP or as a proportion of the modal set.

4 Individual modals by text category over time

So far we have examined change in modal frequency over time by considering two different variables independently: the spoken 'genre' or text type, and the particular core modal used. In this section we investigate the interaction between these two variables: whether particular core modals change at different rates depending on the text type. This implies a more complex experimental design, which we discuss below.

As in section 3, we analyse changes in the rate of individual core modal use for different text types as a proportion of tensed verb phrases (tVPs), and as a proportion of core modals. This allows us to differentiate what we referred to as 'typical' and 'atypical' change. For simplicity we will record changes in terms of percentage swing (rather than simple swing) where these are significant.

The first thing we can do is try to summarise significant change over time in a table. Table 6 colour-codes percentage change (black = fall, grey = rise, 'ns' = non-significant) for each cell against a tensed VP baseline. In this table the modals have been reordered to place potential alternates (e.g. *can*/*may*, *shall*/*will*) side by side for ease of comparison. We know that modals are falling in use generally (final column), which explains the preponderance of black cells expressing negative change, but it should also be immediately clear that actual changes differ according to the particular verb and spoken text type.

	can	*may*	*could*	*might*	*shall*	*will*	*should*	*would*	*must*	**All**
formal face-to-face	ns	ns	ns	ns	ns	ns	-60.04%	ns	-74.85%	ns
informal face-to-face	27.19%	-42.48%	ns	47.34%	-31.54%	ns	ns	ns	-52.84%	ns
telephone conversations	-36.50%	ns	-44.17%	ns	-56.04%	-29.50%	ns	-44.36%	ns	-35.11%
broadcast discussions	-41.20%	-58.63%	ns	ns	-83.31%	ns	ns	ns	-54.12%	-20.19%
broadcast interviews	ns	-60.69%	ns	-58.51%	ns	-41.49%	-55.32%	-31.97%	-56.97%	-35.18%
spontaneous commentary	ns	ns	ns	ns	-92.60%	57.97%	ns	ns	-64.47%	ns
parliamentary language	ns	ns	ns	ns	ns	-39.38%	ns	-29.83%	ns	-20.22%
legal cross-examination	304.37%	ns	ns	ns	ns	ns	1,264.75%	253.83%	ns	157.25%
assorted spontaneous	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
prepared speech	ns	-63.36%	ns	ns	ns	327.21%	ns	-31.73%	-47.65%	ns
All genres	ns	-39.77%	-11.18%	ns	-48.31%	12.94%	-13.61%	-6.98%	-54.19%	-6.18%

Table 6: Significant changes in the proportion of individual core modals out of tensed verb phrases, expressed as a percentage swing from LLC (black = fall, grey = rise, 'ns' = non-significant at $p<0.05$).

Table 6 views change against tensed VP frequency, where each tVP could conceivably include a core modal. However, as we have seen, changes within the set of core modals are inevitably obscured by the overall decline in modals. We examined how text categories differ in changes in frequency of all modals in section 2 (see Figure 9 in particular). The column marked 'All' summarises these changes, where they are significant. Thus we know that in telephone conversations and broadcast interviews we are seeing a significant fall of around 35%, whereas broadcast discussions and parliamentary language have around 20% fewer core modals. (We also saw a major increase in the use of modals in legal cross-examination texts.) On the other hand, the two face-to-face conversational categories do not provide evidence of significant change.

Table 7 focuses on individual changes *within the set of core modals* and factors out the effect of this modal-set variation. As in section 3, we employ the total number of core modals as a baseline. This allows us to more easily contrast adjacent columns, such as *can* and *may*, and *shall* and *will*. Table 7 reveals something else in comparison with Table 6: by only examining change within the core modal set we discount some variation, such as that seen in legal cross-examination texts, where few participants are found, and where variation may be attributable to individual speaker 'styles'.[18]

[18] Note that even with a focus on within-set variation in the core modals, this paper is not an *alternation study* (see, e.g. Aarts, Close and Wallis forthcoming 2012), which would consider whether one modal was being replaced by a semantically related form. Rather, the present paper can be characterised as a *survey of the pattern of change* over time in a corpus. The fact that *can* significantly increases by 25% per modal set in the informal face-to-face category whereas *may* significantly declines by 43%, for example, does not mean that we can say with any certainty that they are alternating in this category, although this hypothesis would be worth further investigation. Were we to wish to do this we would need to (a) consider *can* and *may* against a baseline of *can+may*, and (b) restrict the data to cases where alternation was feasible. This issue is discussed further in 5.3 below.

	can	*may*	*could*	*might*	*shall*	*will*	*should*	*would*	*must*
formal face-to-face	ns	ns	ns	ns	ns	ns	-56.62%	ns	-72.70%
informal face-to-face	25.15%	-43.40%	ns	44.98%	-32.63%	ns	ns	ns	-53.60%
telephone conversations	ns	ns	ns	ns	ns	ns	ns	ns	ns
broadcast discussions	-26.33%	-48.16%	ns	ns	-79.08%	43.31%	ns	39.88%	-42.51%
broadcast interviews	61.98%	ns	ns	ns	ns	ns	ns	ns	ns
spontaneous commentary	ns	ns	ns	ns	-93.40%	40.95%	ns	ns	-68.30%
parliamentary language	ns	ns	116.25%	ns	ns	ns	115.22%	ns	ns
legal cross-examination	ns	ns	ns	ns	ns	ns	ns	ns	ns
assorted spontaneous	53.93%	ns	ns	ns	ns	ns	ns	ns	ns
prepared speech	-28.94%	-67.31%	ns	ns	ns	281.09%	-43.25%	-39.10%	-53.30%
All genres	13.91%	-35.80%	ns	ns	-44.90%	20.38%	ns	ns	-51.17%

Table 7: Significant changes in relative frequency of use (DCPSE: 1960s to 1990s) of core modals, expressed as a proportion of the 1960s data (black = fall, grey = rise, 'ns' = non-significant at $p<0.05$).

One problem with Table 7 is that it is not easy to objectively determine which columns or rows are the most similar or distant. In other words, we cannot easily spot where core modals behave similarly across text categories, or where text categories appear to be similar in their modal use. Among the core modals, we placed potential alternates side by side, but these are the cases we expect to differ.

A more objective algorithmic approach employs a computational method borrowed from psychology, termed *Repertory Grid Analysis* (RGA; Kelly 1955).[19] RGA takes a grid like Table 7 as input after each difference value is translated into a simple ranked score. In our case, we use a five-point scale, where 1 represents a significant fall and 5 a significant rise in the usage of a particular modal. The centre value 3 is non-significant, and we assign intermediate values to cells that would be significant if we relaxed the error level to $p<0.1$, providing some additional information to improve the matching. We then use an algorithm which creates a similarity score for each pair of rows or columns by adding the differences in ranked scores, and then uses this to pair the most similar rows or columns together repeatedly, creating a binary cluster tree or 'dendrogram'.

This procedure results in two cluster trees, one for modals (Figure 11) and one for text categories (Figure 12). We will discuss each of these figures in turn. They are best interpreted in combination with the grid in Table 8, which has been reorganised using the clustering information: the modals and text categories have been reordered to match the ordering in the cluster trees, and the dotted lines indicate major groups identifiable in the trees.

This means that the table now groups together those modals which show more similar patterns of change across each of the text categories, and those text categories which show more similar patterns of change across each of the modals. Table 8 also includes an indication of the intermediate change values used in the RGA analysis, shown as shaded boxes marked 'ns' (non-significant at $p<0.05$). If you compare Tables 7 and 8 you can see, for instance, how this reorganisation has placed most of the decline in the share of modals in the lower right quarter of the table.

[19] We are concerned here not with Kelly's psychological theory, but with the algorithm for clustering similar patterns.

	could	*should*	*might*	*would*	*will*	*shall*	*must*	*may*	*can*
broadcast interviews	ns	ns	ns	ns	ns	ns	ns	ns	61.98%
assorted spontaneous	ns	ns	ns	ns	ns	ns	ns	ns	53.93%
telephone conversations	ns	ns	ns	ns	ns	ns	ns	ns	ns
legal cross-examination	ns	ns	ns	ns	ns	ns	ns	ns	ns
parliamentary language	116.25%	115.22%	ns	ns	ns	ns	ns	ns	ns
formal face-to-face	ns	-56.62%	ns	ns	ns	ns	-72.70%	ns	ns
spontaneous commentary	ns	ns	ns	ns	40.95%	-93.40%	-68.30%	ns	ns
broadcast discussions	ns	ns	ns	39.88%	43.31%	-79.08%	-42.51%	-48.16%	-26.33%
prepared speech	ns	-43.25%	ns	-39.10%	281.09%	ns	-53.30%	-67.31%	-28.94%
informal face-to-face	ns	ns	44.98%	ns	ns	-32.63%	-53.60%	-43.40%	25.15%

Table 8: Result of the Repertory Grid Analysis of Table 7 clustering core modals and text categories, employing intermediate change values (shaded, but marked 'ns', with dark grey showing a fall and light grey a rise). Dotted lines represent major groups identified by the algorithm.

The power of the cluster analysis can be seen when we examine the cluster trees in more detail. Consider first the modals cluster tree in Figure 11. The algorithm identifies the patterns of change of *could* and *should*, and *shall* and *must*, as being the most similar, and so on up the tree. Conversely, we can say that the set {*shall*, *must*, *may*, *can*} are distinct from the remaining core modals in their patterns of change (and within this group, *can* differs from the rest, as the heavier line indicates). Table 8 illustrates what this means in practice: members of this set (grouped together on the lower right) significantly decline in number over several categories.

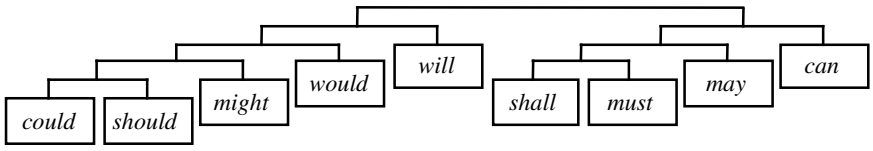

Figure 11: RGA cluster tree of relative (within-modal) changes over time: modals are paired for their similarity of rises and falls over text categories. The most similar terms are grouped lower down the tree. Heavy lines indicate greater differences (similarity score > 10).

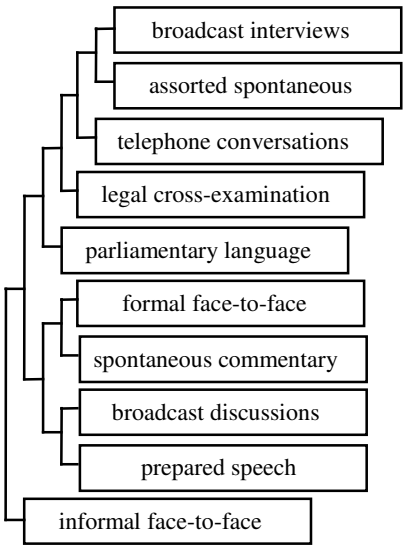

Figure 12: RGA cluster tree of text categories, grouped by similarity of diachronic within-modal change.

Consider next the cluster tree of text categories in Figure 12. This reveals two groups and one outlier: informal face-to-face conversation. A glance at Table 8 reveals that the upper group is mostly an area of non-significant change. The second group – formal conversation, spontaneous

commentary, broadcast discussions and prepared speech – is where most of the changes over time are to be seen, and their patterns of change are mostly distinct. Finally, informal face-to-face conversation, the largest category in the corpus (see Table 1), exhibits a different pattern again.

The cluster tree for the text categories does not suggest any obvious sociolinguistic explanations for the groupings. For example, the three categories which consist mainly of monologue – assorted spontaneous speech, spontaneous commentary and prepared speech – do not form a cluster. Nor do we see separate groups of public and private categories (where 'public' categories are taken as those with an audience other than the immediate participants). The outlying category of informal face-to-face conversation could be said to represent one end of a spectrum, with other categories distinct from it along one or more of the dimensions of formal vs. informal, public vs. private, or monologue vs. dialogue. The telephone conversation category is not so clearly distinct in this regard, although it does seem to differ from informal conversation in including some material from formal contexts alongside that from informal contexts.

The text categories in the upper group, where few significant changes are found, are also the categories which are smaller in terms of word counts. For the most part they are also those with the lower raw frequency counts of modals.

A further visualisation of the data is provided in Figure 13. This graph allows us to show the confidence intervals for each change, which have not so far been presented. Here we limit the range to ±100% and plot 95% confidence intervals as 'I'-shaped error bars. The text categories and modals are displayed in the same order as in the cluster trees and Table 8, with the modals shown in two columns which match the two major groupings revealed by the cluster analysis. As with the reorganised Table 8, these orderings more clearly reveal the patterns in the data.

5 Discussion

In the course of this study, we have examined changing modal usage against several different baselines, moving from the general to the more specific: from words, to tensed VPs, to the set of core modals (or modal tensed VPs). We have shown that the move from words to tensed VPs allows us to eliminate confounding variation in tensed VP density across time and text category. We therefore reported results against a baseline of tensed VPs. However, we then saw that patterns of change for individual modals were obscured by the overall pattern of modal decline. Our final step was to examine these changes against a modal-set baseline. Repertory grid cluster analysis was used as a way to summarise the major patterns in this final data.

In the following subsections we summarise our findings, first from the perspective of text category (5.1) and then from the perspective of groups of modals which show somewhat similar behaviour (5.2). We then discuss the issue of possible alternations between forms (5.3).

5.1 Findings considered by text category

We first consider our results from the perspective of text category. We summarise by noting for each text category, first, *the overall trend* for all modals per tensed VP (Table 6, final column; also Table 4 and Figure 9), and second, *which individual modals* increase or decrease as a proportion of the set of core modals (Tables 7 and 8, and Figure 13). The second type of variation concerns individual modals which behave 'atypically' with regard to the overall trend for each category (variation of the first type).

In the largest category, informal face-to-face conversation, total modal frequencies are stable. The pattern of individual modal change in this category is revealed by the cluster analysis as identifying an 'outlier', a category showing different patterns of change from the others. Three modals (*must*, *may* and *shall*) decrease their share of the modal set, while two (*might* and *can*) increase their share.

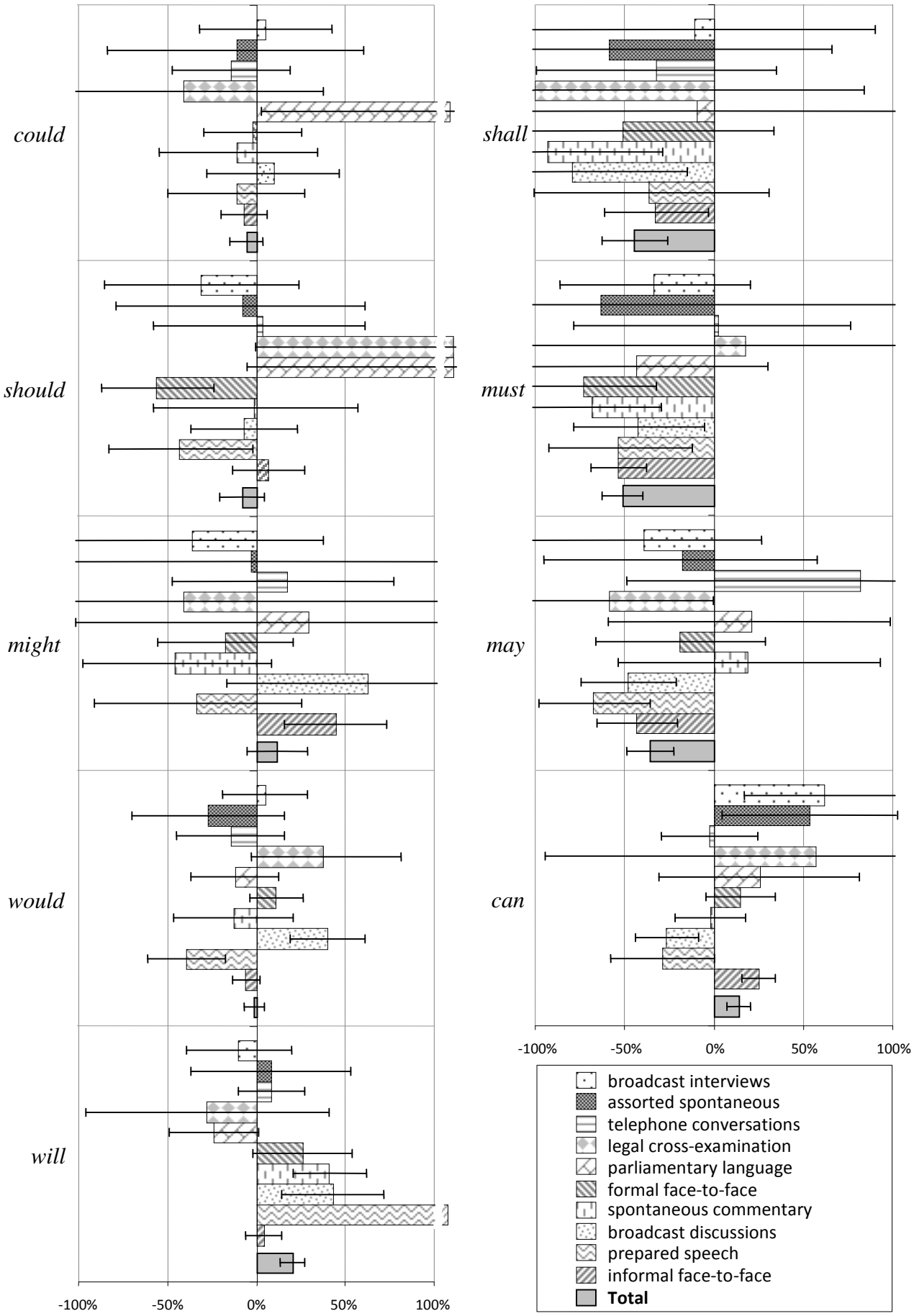

Figure 13: Percentage swing within the core modal set (1960s to 1990s) by text category, with 'I'-shaped 95% confidence intervals. Where an interval crosses the zero axis, the change is not significant.

Most of the remaining changes for individual modals are seen in a group of four text categories: formal face-to-face conversation, spontaneous commentary, broadcast discussions and prepared speech. With one exception, each of these categories shows stable *total* modal frequencies per tVP. The exception is broadcast discussions, where modal frequency declines by 20% (±8.5%). In terms of individual modals, *must* decreases its share of the modal set in all four of these categories. The remaining significant within-set changes are as follows:

- In formal conversations, *should* decreases its proportion of the set.
- In spontaneous commentary, *shall* decreases; whereas *will* increases its share.
- In broadcast discussion, *may*, *shall* and *can* decrease; *will* and *would* increase their shares.
- In prepared speech, *may*, *would*, *can* and *should* decrease; while *will* increases its share.

The remaining five text categories show far fewer significant changes for individual modals, although several categories show changes in total modal frequencies:

- In broadcast interviews, total modal frequencies decline by 35% (±11%), and *can* increases its share of the modal set.
- In assorted spontaneous speech, total modal frequencies are stable, but *can* again increases its share of the modal set. However, as only 13 speakers are represented in this category, the results should be treated with some caution.
- In parliamentary language, total modal frequencies decline by 20% (±14%), and *could* and *should* increase their shares of the modal set.
- In telephone conversations, total modals decrease by 35% (±9%), with no modals behaving 'atypically' (i.e. they appear not to change in proportion to each other).
- The final category, legal cross-examination, is alone among the ten text categories in reporting a significant increase in total modal frequencies. It has no 'atypical' modals. However, as mentioned earlier, this is the smallest text category in the corpus (with a total of 9,338 words, representing only 9 speakers); the total modal frequency is also low at 151. Therefore the results should be interpreted with caution.

5.2 Findings for modals considered in groups

To simplify our summary, we group together modals which show partially similar patterns. We focus here on statistically significant patterns of change within the modal set. Findings for all text categories taken together are compared with data from Leech *et al*. (2009) for written British English in the LOB and FLOB corpora, as reanalysed against a modal-set baseline by Aarts, Wallis and Bowie (forthcoming 2012).

5.2.1 The modals *must*, *shall* and *may*

The modals *must*, *shall* and *may* are the only ones which decline as a proportion of the core modal set when all text categories are considered together. They show substantial declines: 51% (±11%) for *must*, 45% (±19%) for *shall*, and 36% (±13%) for *may*. In our other study mentioned above (Aarts, Wallis and Bowie forthcoming 2012), we performed a within-set analysis of the same core modals in the British written English corpora LOB/FLOB and compared these with DCPSE. In LOB/FLOB the comparable figures are declines of 22% (±8%) for *must*, 38% (±13%) for *shall* and 9% (±7.5%) for *may*. Moreover, the results for *must* and *may* were found to be statistically separable. That is, the relative declines for those two modals are significantly greater in the spoken corpus, suggesting that spoken English is 'leading the way' for these changes.

It is notable that *must*, *shall* and *may* are low-frequency modals (cf. Table 5 and Figure 10). Leech *et al.* (2009: 90, cf. also 269–70), comparing frequency changes in individual modals and in particular uses of modals, suggest 'a general principle that "losers lose out" (the infrequent suffer loss more than the frequent do)'. This idea draws on arguments by functional and cognitive linguists

that greater usage strengthens cognitive representations (e.g. Evans and Green 2006: 114, 118–20). It seems likely that this principle is implicated in the patterns of change we observe. However, it cannot fully explain them: for one thing, other low-frequency modals such as *might* and *should* do not share this pattern of steep decline.

Must, *shall* and *may* are also identified as similar in the RGA tree, showing that a comparable pattern is observed across several individual text categories. They all decline per modal set in the largest category of informal conversations, and they are the only modals to do so. They all also decline per modal set in broadcast discussions. In addition, *may* and *must* fall per modal in prepared speech; *must* and *shall* in spontaneous commentary; and *must* in formal conversation. Figure 13 reveals that in every single text category without exception, *shall* declines in its share of the modals in numerical terms, although this is not significant in some categories.

These three modals thus behave quite consistently across text categories. They show no significant increase in any text category, and each shows a significant decline in several text categories. With their steep overall falls in number, they appear to be 'leading the trend' of modal decline.

5.2.2 The modals *could*, *should*, *might* and *would*

The four items *could*, *should*, *might* and *would* are usually considered to be the past-tense counterparts of present-tense *can*, *shall*, *may* and *will*, respectively. However, the relationships are not straightforward, as noted, for example, in the recent corpus study by Collins (2009): *should* has various specialised uses with no counterpart in *shall* (pp. 48–51); *may* and *might* now seem to function as separate lexemes for many speakers (p. 117–18); and hypothetical uses are more frequent than past time uses for *would* (p. 139) and *could* (p. 108).

In our data, none of these modals changes its share of the modal set when all text categories are taken together. Although they are similar with regard to this overall diachronic pattern, it should be borne in mind that their frequency profiles are distinct. *Would* is the highest-frequency modal in our data, surpassing even *can* and *will*: *would* comprises 25% of all modals, as against 11% contributed by *could*, 6% by *should* and 4% by *might* (Figure 10). The reanalysed data for the written British English corpora shows a different pattern for *could* from that in DCPSE: it increases its share of the modal set by 12.5% (±6.5%), a result which is statistically separable from that for DCPSE (see also Aarts, Wallis and Bowie forthcoming 2012).

These four modals group together in the cluster analysis, although there are differences in their patterns across text categories. *Could* changes its share of the modal set in only one category, parliamentary language, where it shows an increase. *Should* also increases its share in that category; however, it decreases its share in two categories, formal conversation and prepared speech. *Would* also decreases its share in prepared speech, but increases its share in broadcast discussions. Finally, *might* increases its share in the largest category of informal conversations, but otherwise shows no significant changes.

5.2.3 The modals *will* and *can*

The items *will* and *can* stand out as the two modals which *increase* significantly as a proportion of the modal set – *will* by 20% (±7%), and *can* by 14% (±6.5%). Both are high-frequency items (although *would* is of slightly higher frequency in our data). These two modals also increase their shares of the modal set in the reanalysed data for the written British English corpora, where the figures are 6% (±5%) for *will* and 14% (±7%) for *can*. The proportional increase for *will* is significantly greater in the spoken data than in the written (Aarts, Wallis and Bowie forthcoming 2012).

However, in terms of patterns across text categories in DCPSE, the two modals behave quite differently from each other, as their separation in the cluster analysis indicates. *Will* increases its share of the modal set in three categories: spontaneous commentary, broadcast discussions, and

prepared speech. It does not decrease its share in any category. *Can*, by contrast, decreases its share in broadcast discussions and prepared speech, showing an inverse pattern to *will* in these two categories. It increases its share in three categories where *will*'s share seems stable: broadcast interviews, assorted spontaneous, and the large category of informal conversation.

5.3 Patterns of alternation

This study is *not* an alternation study. In an alternation study, one would consider whether one modal was being replaced by a semantically related form, whether another modal or a quasi-modal form. For example, Close and Aarts (2010) find that the quasi-modal *have to* shows a significant increase over time in DCPSE in contexts of possible alternation with *must*, supporting the idea of a replacement process (though only for root meanings; both epistemic and root *must* are found to decline significantly, but epistemic *have to* remains rare).

Table 7, which displays results per modal, is, therefore, at best indicative. It allows us to identify areas where further investigation of alternation would be worthwhile. We placed potentially alternating modals in adjacent columns deliberately to try to draw out where a decline in one type may lead to an increase in another. First, consider *can* and *may*, both 'possibility' modals, which have a complex relationship of semantic overlap and distinction (see, e.g. Collins 2009). They show inverse patterns for all genres together and for the largest category of informal face-to-face conversation: a significant increase in the share of *can*, and a significant decrease in the share of *may*. (We can further note that the possibility modal *might* increases its share in the informal category.) This suggests there may be some degree of replacement of *may* by *can*, a possibility which is worthy of further investigation. On the other hand, in two categories (broadcast discussions and prepared speech) both *can* and *may* decrease their share.

In addition, we can compare the per modal results for *shall* and *will*. These show an inverse pattern both overall, and in the two genres of broadcast discussions and spontaneous commentary: significant and substantial declines in *shall* and increases in *will*. Again, this suggests there may be a trend for *will* to replace *shall* in some contexts. In fact, a detailed alternation study of these two modals in DCPSE (Aarts, Close and Wallis forthcoming 2012) has provided clear evidence of a shift in use from epistemic *shall* to *will* in first-person declarative contexts.

6 Conclusion

The overall trend of modal use in spoken British English over the period 1960–1990 is downwards. This trend is significant and can be identified against a number of baselines, including the common per-million-word baseline. However, this does not mean that modals are all declining in use at the same rate, or even that all core modals are declining in use! Nor does it mean that the same trend that might be found when analysing a broad range of spoken transcripts is found in subcategories of these transcripts.

The first conclusion one might make, therefore, is that different types of texts may be undergoing different changes, and that the assumption that a particular type of text, such as written material, is 'representative' of current change may well be flawed. We have seen that the frequency of some individual modals appears to be changing more rapidly in DCPSE than in the LOB/FLOB pair. Were we to assume that LOB/FLOB 'represented' English of the time period concerned for this change, therefore, we would be misled, at least as far as changes in core modal frequency were concerned. Our study highlights the importance of taking care in corpus sampling (with an ideal of a 'balanced' and 'representative' corpus in mind) and in experimental design (to make the best use of this corpus).

In order to explore the more complex picture which results, we demonstrated that it is necessary to first refine the baseline from one which is proportionate to the number of words in texts to one based on tensed verb phrases. The argument goes that only such verb phrases could conceivably contain a modal verb. This procedure draws out differences between text categories

which would otherwise be missed or conflated with varying ratios of tensed VPs to words, a ratio which for our purposes constitutes 'noise'.

This experimental design can also be modified to consider relative variation within the modal set. Here we are not claiming that the set of modals represents an alternation set, but merely a repertoire of available modal forms. This procedure allows us to differentiate between two types of change: absolute use of the modals by text category, discussed in section 2, and change in the relative share of the modal set taken by a particular core modal.

The patterns in the various spoken text categories present a complex picture. There is considerable variation among text categories synchronically and diachronically. Nevertheless, some diachronic trends emerge quite clearly from this complex picture, and are broadly in accord with those found by Leech *et al*. (2009) and Aarts, Wallis and Bowie (forthcoming 2012) for the core modals in written corpora of BrE and AmE. We used a repertory grid method to help us summarise this data, but the clustering of texts did not suggest a 'neat' explanation of an underlying factor, apart from the fact that it identified a group of texts as showing little significant within-set change. The remaining larger text categories appeared to present quite different patterns of relative modal change.

Millar (2009) questions the robustness of results based on comparing two particular years (or periods) as 'points' in time, as done in work by Leech *et al*. (2009) using the Brown family of corpora and in the present study. Commenting on his data on modal frequencies from the *Time* Magazine corpus from 1923 to the present day, he notes that there is some year-to-year variation in frequencies that does not reflect the longer-term trends he found in his data. This result is not surprising. Unless we are looking at the spread of neologisms from their invention, it would be unlikely for us to be able to plot serial incremental change year on year! Within a closed set such as the core modals, we would expect to find considerable 'noise', with frequencies increasing and decreasing annually. In order to compensate for this we employ sound statistical methods which estimate the likely error margins of change. For more on this see Aarts, Close and Wallis (forthcoming 2012).

We have been careful to recognise the limitations of our own data. In our analysis we noted that some sub-categories of text are relatively small (particularly legal cross-examination) and have few participants. Here we must avoid overstating the case for particular changes. In corpus linguistics we make two important assumptions to permit statistical inference to be employed. We assume that a sample is representative of the data from which it is drawn. However, if we rely on a small number of continuous-text samples, this claim may be difficult to justify. We also assume that every time a speaker makes a decision to employ a particular expression within a text they do so independently from previous decisions. Although this is an idealisation, it is less problematic for large samples made up of many texts. Indeed, the significant within-modal results outlined in the discussion primarily concern the larger text categories with many participants (see Table 1), where such objections are minimised.

One way that the present study could be improved would be to extend the timescale. Obviously it is desirable to have multiple time points or periods for comparison where possible. The trends found in Leech *et al*. (2009) are supported by further data presented in Leech (2011), which extends the picture both forwards and backwards in time. Ideally we would wish to extend DCPSE in a similar way, although the accurate transcription and reliable parsing of speech data are costly to undertake.

Appendix 1: Raw frequency data for modals in ICE-GB by broad genre groupings

lemma	dialogue	non-print	print	monologue	all genres
would	1,682	353	791	578	3,404
will	1,432	454	1,116	859	3,861
can	1,539	378	900	748	3,565
could	693	137	466	326	1,622
may	196	142	677	203	1,218
should	400	106	392	226	1,124
must	199	56	303	117	675
might	314	52	210	126	702
shall	92	30	62	39	223
total	6,547	1,708	4,917	3,222	16,394
all words	360,000	100,000	300,000	240,000	1,000,000

These data are collated from Collins (2009).

Appendix 2: Raw frequency data for modals by text category, 1960s and 1990s subcorpora of DCPSE

1960s	*can*	*may*	*could*	*might*	*shall*	*will*	*should*	*would*	*must*	**Total**	**tVP**
formal face-to-face	187	35	99	53	12	105	60	269	39	859	5,505
informal face-to-face	796	125	428	122	83	669	198	968	241	3,630	27,917
telephone conversations	132	14	91	35	26	227	36	105	23	689	3,226
broadcast discussions	196	92	65	22	19	110	88	179	53	824	5,282
broadcast interviews	49	19	57	16	10	78	26	111	27	393	2,267
spontaneous commentary	122	14	31	19	12	122	20	50	30	420	4,049
parliamentary language	27	15	11	7	5	82	10	85	14	256	920
legal cross-examination	3	10	7	2	2	9	1	21	1	56	740
assorted spontaneous	37	16	18	6	7	39	19	38	4	184	1,310
prepared speech	69	50	46	20	16	52	36	104	35	428	3,275
All genres	1,618	390	853	302	192	1,493	494	1,930	467	7,739	54,491

1990s	*can*	*may*	*could*	*might*	*shall*	*will*	*should*	*would*	*must*	**Total**	**tVP**
formal face-to-face	182	24	82	37	5	112	22	253	9	726	5,051
informal face-to-face	873	62	350	155	49	611	185	798	98	3,181	24,073
telephone conversations	66	13	40	21	9	126	19	46	12	352	2,540
broadcast discussions	109	36	54	27	3	119	62	189	23	622	4,996
broadcast interviews	62	9	47	8	7	55	14	91	14	307	2,732
spontaneous commentary	151	21	35	13	1	217	25	55	12	530	4,559
parliamentary language	30	16	21	8	4	55	19	66	7	226	1,018
legal cross-examination	8	7	7	2	0	11	9	49	2	95	488
assorted spontaneous	39	9	11	4	2	29	12	19	1	126	1,095
prepared speech	48	16	40	13	10	194	20	62	16	419	2,860
All genres	1,568	213	687	288	90	1,529	387	1,628	194	6,584	49,412

The table above summarises raw frequencies in each spoken subcategory or genre, from left to right, for each core modal, the total number of core modals and tensed VPs (tVP column).

References

Aarts, Bas, Joanne Close and Sean Wallis (forthcoming 2012). Choices over time: methodological issues in current change. In: Bas Aarts, Joanne Close, Geoffrey Leech and Sean Wallis (eds.), *The verb phrase in English: investigating recent language change with corpora*. Cambridge: Cambridge University Press.

Aarts, Bas, Gerald Nelson and Sean Wallis (1998). Using Fuzzy Tree Fragments to explore English grammar. *English Today* 14, 52–6.

Aarts, Bas, Sean Wallis and Jill Bowie (forthcoming 2012). Profiling the English verb phrase over time: modal patterns. In: Merja Kytö, Irma Taavitsainen, Claudia Claridge and Jeremy Smith (eds.), *Corpus linguistics and the development 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.

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: Ursula Lenker, Judith Huber and Robert 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, 165–81.

Collins, Peter (2009). *Modals and quasi-modals in English*. Amsterdam; New York: Rodopi.

Davies, Mark (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14, 159–90.

Davies, Mark (2010). More than a peephole: using large and diverse online corpora. *International Journal of Corpus Linguistics* 15, 405–11.

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

Kelly, George (1955). *The psychology of personal constructs.* New York: Norton.

Leech, Geoffrey (2003). Modality on the move: the English modal auxiliaries 1961–1992. In: Roberta Facchinetti, Manfred Krug and Frank Palmer (eds.), *Modality in contemporary English*. Berlin; New York: Mouton de Gruyter, 223–40.

Leech, Geoffrey (2011). The modals ARE declining: reply to Neil Millar's 'Modal verbs in TIME: frequency changes 1923–2006'. *International Journal of Corpus Linguistics* 16(4), 547–64.

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

Millar, Neil. 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.

Newcombe, Robert 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.

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