

Close, Joanne and Bas Aarts. In prep. “Current Change in the Modal System of English: A Case Study of *Must*, *Have To* and *Have Got To*.”

Do not quote without permission.

Version 1 Jan 2009.

Current Change in the Modal System of English: A Case Study of *Must*, *Have To* and *Have Got To*

JOANNE CLOSE & BAS AARTS

UCL

This paper takes the variation between *must*, *have to* and *have got to* as a window through which to view changes in the modal system in present-day British English (1960s-1990s). The results from this study show a dramatic decrease in frequency of the core modal *must* and a significant increase in the semi-modal *have to* in the *Diachronic Corpus of Present-Day Spoken English* (DCPSE). Changes in the modal system affect both epistemic and root uses of *must*, although *have to* is only an active rival to root *must*; epistemic instances of *have to* (and *have got to*) are rare in the corpus.

1. Introduction

The class of English modal verbs have undergone a number of important changes in the history of English, and continue to change in Present-Day English (PDE). Studies on current change suggest that the modals are decreasing in frequency, becoming monosemous, and are facing possible competition from rival semi-modals which are increasing in frequency (see Krug 2000; Leech 2003; Smith 2003; Mair and Leech 2006; Leech *et al.* forthcoming).

This paper investigates the variation between the core modal *must* and the semi-modals *have to* and *have got to* and in the *Diachronic Corpus of Present-Day Spoken English* (see section 3.1), with the aim of answering the following questions: (i) is the core modal *must* decreasing in frequency? (ii) is there evidence for

“competition” between *must* and the semi-modals? ¹ (iii) is *must* becoming monosemous? And (iv) what is the motivation for the change?

The variation between the three forms is a window onto the changing modal system and, while the results cannot necessarily be generalised to other modals, the study will offer an insight into areas of the modal system susceptible to change and identify possible motivating factors behind these changes.

This paper is organised as follows: section 2 reviews some of the existing work on current change in the modal system; section 3 presents the corpus and describes the data collection method, including decisions taken on which data to exclude from the current study; section 4 discusses the main findings and investigates possible motivations for the changes that have occurred; section 5 is the conclusion.

2. *Current change in the modal system of English*

2.1 *Observing current change*

Throughout the history of English, the modal verbs have undergone a number of significant changes, resulting in a class of verbs characterised by a number of syntactic properties which are not shared by main verbs (see Traugott 1989, Lightfoot 1991, Warner 1993). More recently, there has been a growing interest in current change in the modal system of English (Krug 2000, Leech 2003, Smith 2003, Leech *et al.* forthcoming), as well as in current change more generally. It is hoped that research into current change will lead to a greater understanding of the trigger(s) and pathway(s) of language change.

We might expect instances of current change to be subtle in comparison to changes measured over a longer period. While recent changes in the language may not appear as dramatic as some of the earlier changes such as the development of the English inflectional system from Old to Middle English, the shift from OV to VO or the appearance of *do* support, the end result could be equally significant. There exist instances of historical language changes which involve a change in the underlying grammar of the language (see Kroch 1989, Lightfoot 1991, Warner 1993, Pintzuk 1995, among others). Current change has, as yet, not been linked to grammar change in this way. According to Leech (2003:223), however, “if we understand such

[grammatical] changes to include changes of frequency, significant grammatical changes do take place within a generation.”

The nature of current change is such that it is more difficult to observe than a change which has occurred over a longer period of time. In the study of a long-term change, the end-point of the change has typically been reached, and it is often possible to track the change through the period in which it occurred. A change measured over a shorter period of thirty years or so, however, will not necessarily be completed in the time period, and it is impossible to know *how*, *when*, or even *if* a particular change will complete. The danger of attempting to predict these is illustrated by the subjunctive which was believed to be decreasing to the point of extinction (Fowler 1965), but has been shown more recently to be undergoing a revival (Övergaard 1995, Hundt 1998, Leech *et al.* forthcoming). It is, of course, possible to track a change as it occurs, but it will often be necessary to observe the phenomenon at a later period (see Bauer 2002, Mair 2008).

2.2 Core modal frequency in Present-day English

One of the observed changes in the modal system of PDE is a decrease in frequency of the core modals (see Krug 2000, Leech 2003, Smith 2003, Leech *et al.* forthcoming). Leech (2003) investigates modal usage in the Brown quartet of corpora,² and observes that the decrease in modal usage is apparent in 13 out of 15 categories ranging from a decline of 31.2% (D, religion) to a decline of 0.5% (H, Miscellaneous). In the remaining two categories, there is a small and non-significant increase; in J (Learned) of 2.4% and L (Mystery Fiction) of 8.4%. Furthermore, the British and American corpora show a decrease over time (from 1961 to 1991) in the case of every modal (individual modals differ in frequency), except *can* and *could* which show an increase of 2.2% and 2.4%, respectively, in BrE. Leech's results also show that the infrequent modals *shall*, *ought to*, and (in BrE) *need* have decreased drastically, and a similar decline has taken place in the mid-frequency modals *may* and *must*; while *would* and *should* have decreased less drastically. Finally, *will*, *can*, *could* and *might* have not significantly changed. The results presented by Leech *et al.* (forthcoming) also confirm that the modals which show the steepest decline are the less common modals.

2.3 *Core modals vs. semi-modals*

Mair and Leech (2006) point out that although many changes involve a competing construction or constructions, there is not necessarily a correlation between the form(s) losing ground and those gaining. On discovering a decrease in the frequency of the core modals in the Brown family of corpora, Leech (2003: 229) suggests that perhaps the semi-modals are “gradually usurping” the functions of the core modals, but concludes that this is not the case because, although semi-modal usage is increasing overall, some semi-modals are actually declining (for example, *be to*, *(had) better*, *(have) got to* and *be going to* in FLOB), and most semi-modals are much less frequent than the core modals (for example, *will* is about ten times more frequent than *be going to*). These results are supported by Krug (2000), and also by Smith (2003: 249) who states that “the rise of [*have to*] ... by no means makes up for the shortfall in *must*.”

2.4 *Monosemy of modals*

Leech (2003) carries out a semantic analysis of *should* and every third instance of *may* and *must* found in the Brown corpora to discover whether a decline in frequency can be attributed to one particular sense of a modal. He provides some evidence that the decline in frequency of the modals is linked to the tendency for modals to become “more monosemous”, but points out that this is not always the case. In fact, a different pattern can be observed for each of the three modals investigated. *Must*, the modal we are interested in here, shows a decline in both root and epistemic senses between the 1960s and the 1990s, particularly in the SEU-mini and ICE-GB-mini corpora of spoken English,³ but the root (or deontic) sense has shown the largest decrease in frequency.⁴ Leech warns that the results of the study as a whole are tentative because of uncertainties in sampling and coding procedures. Further to Leech’s warning it is important to add that comparing the results of semantic analyses across studies is difficult due to the challenging nature of the task of semantic coding.

3. Corpus and data collection

3.1 *Current change and the Diachronic Corpus of Present-Day Spoken English*

There are a number of “pitfalls of anecdotal observation” as far as current change is concerned, most notably “[t]he spread of salient new uses is exaggerated, while the less salient persistence of older forms is not noted or [...] a diachronic trend is read into a situation which merely shows variable or fluctuating usage” (Mair 2008). For this reason, studies on current change are carried out more reliably using corpora.

One of the strengths of the research reported on here is that it uses a corpus of spoken English. This is particularly important for the study of the semi-modals as “these forms, being typically colloquial, are not likely to show up in their true colours in the written language” (Leech 2003: 230).

The *Diachronic Corpus of Present-Day Spoken English* (DCPSE) contains 421,362 words from the British component of the *International Corpus of English* (ICE-GB) collected in the early 1990s and 464,074 words from the *London-Lund Corpus* (LLC) collected between the late 1960s and early 1980s (for more information about the LLC and ICE-GB see Svartvik (1990) and Nelson *et al.* (2002), respectively). The period of time between LLC and ICE-GB is within the range of what is considered ideal for studying current change. In designing the Brown family of corpora a decision was made to have an interval of approximately thirty years between the Brown and LOB corpora (1960s) and the Frown and LOB corpora (1990s) because “the interval of one generation or so ... is usually considered the minimum sufficient to clearly identify and document linguistic change in real time” (Mair 1998: 140).

With over 800,000 words of spoken English, DCPSE is the largest available single collection of parsed and checked orthographically transcribed spoken material. All of the sentences in DCPSE have been grammatically analysed and have been given a detailed parse tree. Figure 1 is an example of a tree diagram of a simple sentence involving a subject, verb and direct object. The tree is shown from left-to-right here. Each of the ‘boxes’ are called *nodes*. These have three sections: the top left section contains *functional information* (subject, direct object, noun phrase head, etc.); the top right section contains *categorial information* (part-of-speech, phrase, clause,

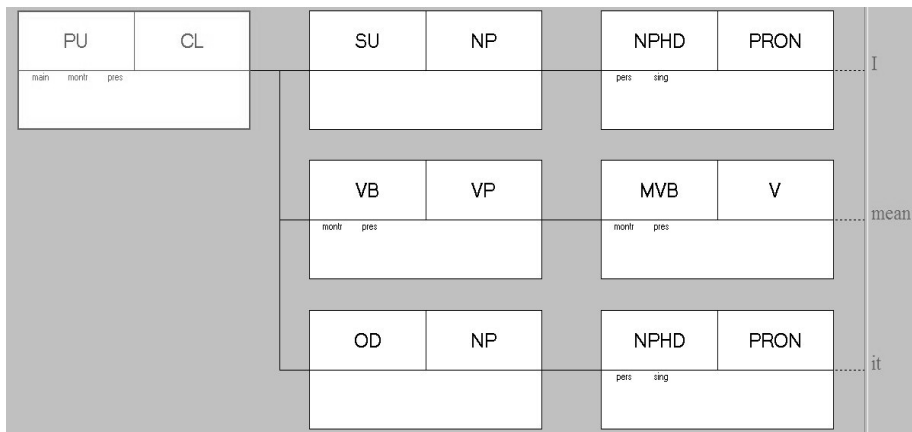


Figure 1: Tree diagram for I mean it.*

*Gloss for figures (lower case items are 'features'): PU=parsing unit, CL=clause, SU=subject, NP=noun phrase, VB=verbal, VP=verb phrase, OD=direct object, NPHD=NP head, PRON=proun, MVB=main verb, V=verb, N=noun, main=main clause, montr=monotransitive, pres=present, pers=personal, sing=singular.

etc.); and the bottom section contains optional additional *features* (intransitive (verb), common (noun), main (clause), etc.).

The corpus can be explored by using the *International Corpus of English Corpus Utility Program* (ICECUP) software developed at the Survey of English Usage. ICECUP offers a facility called *Fuzzy Tree Fragments* (FTFs; see Aarts *et al.* 1998 and Nelson *et al.* 2002).⁵ An example FTF which searches for all noun phrases (NPs) in the corpus is shown in Figure 2.1. This FTF is a partial tree diagram, which can be constructed by users. FTFs are 'fuzzy' in the sense that users can specify for themselves the level of detail that they are interested in. For example, if the user adds 'SU' to the top left section of the FTF in Figure 2.1, the software will search for all noun phrases in the corpus that function as subjects. This is shown in Figure 2.2.

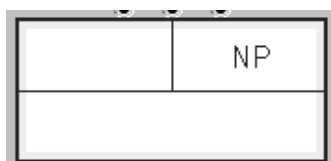


Figure 2.1: An FTF which instructs the search software to find all the noun phrases in the corpus.

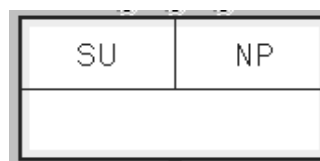


Figure 2.2: An FTF which instructs the search software to find all the noun phrases in the corpus functioning as subjects.

@ @Insert Figures 2.1 and 2.2 here.

@@Insert Figure 3 here.

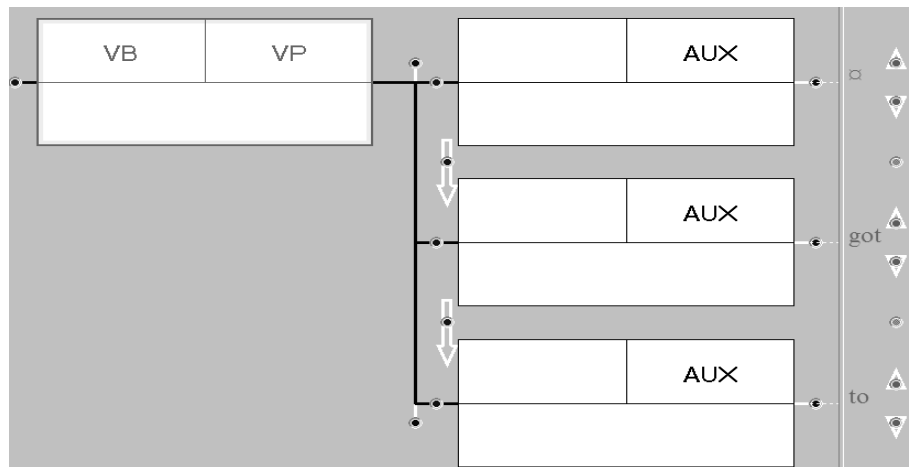


Figure 3. FTF for *(have) got to*.

More complex FTFs can contain grammatical structures, words, lexical wild cards, etc. In the context of the present research, the FTF in Figure 3 searches the corpus for the string *got to* with the position for *have* left unspecified (indicated by the symbol ‘□’). This FTF will find cases of *‘ve/’d got to* as well as *have/has/had got to*, and even cases where an adverb intervenes between *have* and *got*. This is because in Figure 3 the arrows between the nodes have been set in such a way that something may intervene between the nodes. This is an example of user choice. Users can click on the ‘radio buttons’ superimposed on the arrows to change this setting, for example to specify that the nodes should follow each other immediately.

3.2 Data

In order to ensure reliability of results, it is important to identify contexts in which the forms being investigated may alternate. In the case of *have to*, *have got to*, and *must* there are a number of syntactic environments in which the forms are not variable, and these have been excluded from the study (see Tagliamonte and Smith 2006, Depraetere and Verhulst 2008). All exclusions are listed below, and are not included in the statistical tests in section 4.

First, negated forms were excluded because the meanings are not equivalent:⁶

- (1) a. You mustn’t go (the modal is within the scope of the negative: ‘You are not allowed to go’).

- b. You don't have to go (the modal has scope over the negative: 'You are allowed to not go.').
- c. You haven't got to go (the modal has scope over the negative: 'You are allowed to not go.').

Secondly, interrogative contexts were excluded because interrogative forms with *must* did not occur at all in the later period, and only one example occurred in the early period:

- (2) Must Wordsworth speak for us in his intimations of immortality?
(DCPSE:DL-J01 #0129:2:B)

Finally, also excluded were non-finite *have to* (including future forms) because equivalent non-finite forms of *must* or *have got to* do not exist; and the past form *had to* as *must* cannot refer to past tense.

It is important to point out that tokens of *must have been* or *must have V-ed* have not been excluded. Although there are no tokens of *have (got) to have been* or *have (got) to have V-ed* in the corpus, a search of the web using WebCorp illustrates that both of these forms are possible.⁷ An example of *have to have been* is given in (3):

- (3) Just my tuppence worth, but 2007 has to have been the absolute worst X Factor *ever*.⁸

3.3 *Semantic coding*

A number of claims have been made in the recent literature about the frequency of the modals and semi-modals relating to meaning. For instance, Coates (1983) states that *have to* and *have got to* are infrequent as epistemic modals and Leech (2003) claims that *must* is one of a few modals not becoming monosemous. In order to test these claims using DCPSE, each instance of *must*, *have got to* and *have to* was manually classified according to the type of modality that they express.

Modal meaning has been dealt with in a number of ways. This work uses the coding system advocated by Coates and Leech (1980) and Coates (1983) in which

modals are divided into two semantic classes, Root and Epistemic. In this approach modal meaning is dealt with in a gradience model where each class has both core and peripheral members. This can account for the fuzziness in modal meaning without the need for postulating additional classes. This is described in more detail below.

3.3.1 *Epistemic modality*

Epistemic *must* ‘conveys the speaker’s confidence in the truth of what he is saying, based on a logical process of deduction from facts known to him (which may or may not be specified)’ (Coates 1983: 41). Coates divides epistemic *must* into core (confident inference) and peripheral (logical necessity) meanings:

- (i) *confident inference*: ‘I confidently infer that x’
- (ii) *logical necessity*: ‘In light of what is known, it is necessarily the case that x’

Epistemic examples of *must*, *have got to* and *have to* are given in (4) below:

- (4) a. I am coming to conclude that it must be because I get too much water in it or something. (DCPSE:DI-B47 #0166:1:B)
- b. Loose shirts over jeans has got to be a sort of temporary prejudice. (DCPSE:DI-B44 #0161:1:B)
- c. I mean good guidebooks really are full of impressions aren't they whereas imaginative works or works of fiction often do blend facts in a distorted form if you think of Hardy 's use of landscape as a curious blend of the identifiable which people who research in that find interesting, and something which has to be an impression there wouldn't you agree? (DCPSE:DL-A06 #0310:1:E)

3.3.2 *Root modality*

Coates argues that Root modality is fuzzy with a cline from strong obligation (‘It is imperative/obligatory that x’) to weak obligation (‘It is important that x’), although the basic meaning of Root *must* is ‘It is necessary for...’ Examples of Root meaning are given in (5):

- (5) a. My boss had said you must read those books. (DCPSE:DI-B12 #0171:1:E)
- b. I ve got to go, I'm afraid, in an hour. (DCPSE:DI-B37 #0216:2:B)
- c. Actually you have to be much more honest about what you're doing. (DCPSE:DI-A02 #0124:2:B)

3.3.3 *Performative modality*

Included in the class of Root modality are cases in which the speaker is carrying out the action denoted by the verb. These are known as Performative or Speech Act modals, and occur frequently with verbs like *admit*, *say*, *confess*, as shown in (6).

- (6) a. There's a piece here called Spring Fire which takes my interest, I must say. (DCPSE:DI-D12 #0141:1:A)
- b. Pretty much yeah I ve got to admit. (DCPSE:DI-A07 #0094:1:B)
- c. I have to confess an often irking thought of am I really really two pounds less than Kate Hamilton. (DCPSE:DI-B07 #0224:2:B)

Although Performative modality is considered to be a type of Root modality, these examples were coded and counted separately because of their specialised meaning.⁹

3.3.4 *Ambiguity*

Coates (1983: 47) asserts that “there is no overlap between the two fuzzy sets representing Root and Epistemic *must*. Cases where it is not possible to decide which meaning is intended are therefore ambiguous”, as illustrated in (7) taken from Coates (1983: 47, her example (35)):¹⁰

- (7) And anyway, I think mental health is a very relative thing. I mean, mental health must be related to the sort of general mentality, or whatever other word you use of the community you're living in. (S.2.14.73)

The Root meaning here is ‘it's essential that mental health is related to ...’, while the Epistemic reading is ‘it's inevitably the case that mental health is related to...’.

It might be expected that ambiguous cases would be relatively frequent. However, from the Lancaster and Survey corpora, Coates classes only 11 examples

out of a total of 436 as ambiguous.¹¹ In the ambiguous cases “syntactic features which could disambiguate such [ambiguous] utterances are either not present or are themselves ambiguous.” Coates (1983: 47).

In coding the data from DCPSE, it was assumed that the speaker had an intended meaning when producing the utterance and an attempt was made to classify each example, keeping ambiguous cases to a minimum. The audio files were invaluable here. Despite this, there exist a number of examples which are impossible to disambiguate for two main reasons: (i) the utterance is compatible with both Root and Epistemic meanings and/or (ii) the modal is followed by an ellipsis site or unclear words. The numbers of ambiguous cases are shown in Table 3.

4. *Results and discussion*

4.1 *Overall frequencies*

To examine a possible trend for declining core modal usage and investigate the theory of competition between core modals and semi-modals, the overall frequency of core modal *must* is compared with frequencies of the semi-modal forms in DCPSE. As shown in Table 1, during the thirty-year period the frequency of *must* has declined by over 53% while the frequency of *have to* has significantly increased (+24.27%).

Somewhat surprisingly, *have got to* decreases in frequency by -8.21%. It is impossible for us to be sure whether *have got to* has reached its peak and is showing a decline in the 1990s data prior to “levelling out” in the future, or whether the decrease is simply fluctuating usage in the corpus. To be certain, it will be necessary to measure the frequency of *have got to* at a point later than 1992. Unfortunately, the lack of available corpora makes this impossible at this point in time.

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	χ^2 score
<i>must</i>	427	81.88	172	38.21	-53.33	36.29
<i>have got to</i>	187	40.08	156	36.79	-8.21	3.10
<i>have to</i>	188	40.30	225	50.08	+24.27	31.94
Total	802	162.26	553	125.07	-22.92	71.32

Table 1. Overall frequencies of *must*, *have got to* and *have to* in DCPSE

(figures in bold are significant at $p < 0.01$).

@@ Insert Table 2 here.

Source corpus	Epistemic		Root		Performative		Ambiguous		Total
	N	%	N	%	N	%	N	%	N
LLC	43.96	47.78	36.63	39.81	9.48	10.30	1.94	2.11	92.01
ICE-GB	21.12	51.74	16.14	39.53	2.37	5.81	1.19	2.91	40.82
TOTAL	65.08	48.99	52.77	39.73	11.85	8.92	3.13	2.35	132.83

Table 2. Distribution of semantic types of *must* in DCPSE

(N=frequency per 100,000 words).

4.2 *Root and Epistemic modality*

One of the aims of this study is to determine whether *must* is becoming monosemous. Using the semantic coding carried out on the data (see section 3.3 above, for details), the frequencies of epistemic, root and performative uses were calculated. The results are presented in Table 2.

The results in Table 2 show that both root and epistemic *must* have decreased at a similar rate; epistemic *must* has decreased by -51.96% (from a frequency of 43.96 in LLC to 21.12 in ICE-GB), and root *must* by -55.94% (from a frequency of 36.63 in LLC to 16.14 in ICE-GB). Interestingly, when viewed as proportions of total *must*, the proportion of root *must* remains almost constant (39.81% in LLC and 39.53% in ICE-GB), and epistemic *must* actually shows an increase of just under 4% (from 47.78% in LLC to 51.74% in ICE-GB). These results concur with those presented by Leech (2003): *must* does not appear to be becoming monosemous.

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	χ^2 score
<i>must</i>	174	37.49	70	16.61	-55.69	23.61
<i>have got to</i>	184	39.65	151	35.84	-9.61	0.12
<i>have to</i>	185	39.86	208	49.36	+23.83	12.32
TOTAL	543	117	429	101.81	-12.98	36.05

Table 3. Frequencies of root *must*, *have got to* and *have to* in DCPSE (figures in bold are significant at $p < 0.01$).

To discover if the decline in root and/or epistemic *must* is related to the use of the semi-modals, root and epistemic uses of the three forms were compared. The results for root and epistemic uses are presented in Tables 3 and 4, respectively.

Table 3 illustrates that as expressions of root obligation, *must*, *have got to* and *have to* combined have decreased by -12.98%. Again, the decrease in *must* and the rise in *have to* are found to be significant, and *have got to* shows a decline.

Results for epistemic use in Table 4 show that, although epistemic *must* has decreased by over 50%, *must* is still the preferred form for expressing epistemic necessity; epistemic *have got to* and *have to* are very rare in DCPSE.¹²

The general stance in the literature is that the decline of the core modals cannot be attributed directly to an increase in the semi-modals (Krug 2000; Leech 2003; Smith 2003; Mair and Leech 2006). This conclusion was reached by studying frequencies in written language where, on the whole, core modals are more frequent than semi-modals. The situation with respect to the spoken data is quite different. In root contexts (see Table 4), *have to* is already more frequent than *must* in the 1960s (frequencies are 39.86 for *have to* and 21.60 for *must*). By the 1990s, the frequency of *have to* is almost three times that of *must* (the frequencies are 49.36 and 16.61, respectively). This suggests that there may be a link between modal decline and semi-modal increase.

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	χ^2 score
<i>must</i>	206	44.39	91	21.60	-51.34	0.19
<i>have got to</i>	2	0.43	4	0.95	+120.93	3.36
<i>have to</i>	2	0.43	3	0.71	+65.12	1.83
TOTAL	210	45.25	98	23.26	-48.61	5.38

Table 4. Frequencies of epistemic *must*, *have got to* and *have to* in DCPSE.

4.3 Motivating factors

The frequency shift in the expression of root modality suggests a possible correlation between the decline of *must* and the rise of *have to*. The question, of course, is what is the motivation for the shift away from *must* and towards *have to* in root meaning? A

number of suggestions to explain the decline in *must* have been made in the literature, many of which refer to the “authoritarian” nature of *must*. Myhill (1995), for instance, attributes the decline in *must* (=obligation) and the rise of *should* (=weak obligation) around the time of the American Civil War to a growing tendency to avoid overt claims to authority by the speaker/writer. He claims that the “old” modals were linked with people controlling the actions of other people while the “new” modals were more personal, and used to give advice to an equal, etc. He attributes changes in the modal system to a growing tendency to avoid overt claims to authority by the speaker/writer (a trend he refers to as *democratisation*). Smith (2003) shares a similar view, and makes the following remark:

It seems probable that MUST is a casualty of a changing society where increasing emphasis is being placed on equality of power, or at least the appearance of equality of power, and the informality of discourse found in private conversation is becoming more acceptable, even usual, in official types of discourse. Just as these conditions are likely to disfavour the use of MUST, they should correspondingly favour other forms which express obligation less directly (Smith 2003: 259).

The problem with the suppositions by Myhill and Smith, however, is that epistemic *must* has also decreased, and this verb is not related to power and authority. Leech *et al.* (forthcoming) suggest that the “partial decline [of epistemic *must*] could be due to contamination by the dramatic fall of deontic *must*”. This is not an attractive explanation either because root and epistemic meanings of *must* do not intersect, so it is not clear how a fall in one would cause a fall in the other (cf. Coates’s (1983: 170) treatment of the meanings of *will* which do intersect). In their data, root *must* declines more drastically than epistemic *must*, which Leech *et al.* suggest is because “neither *have to* nor any other form has become widely adopted as an alternative expression of strong epistemic necessity”. In the spoken data from DCPSE, however, the fall of epistemic *must* is almost equal to that of root *must*, although it is true that there is no rise in epistemic *have to* or *have got to*. It is, of course, possible that there are independent explanations for the fall of root *must* on the one hand, and epistemic *must* on the other. However, we suggest that the root cause is the same.

As an explanation for the decrease in *must*, we appeal to what Huddleston and Pullum *et al.* (2002: 175) refer to as “strength of commitment” where “necessity involves a strong commitment”. We suggest that the decline in *must* is a result of a decline in forms expressing strong commitment. In the root sense, *have to* is favoured over *must*. The difference between the two forms is illustrated in (8) where the conjoined clause is not possible with *must*, because the commitment is stronger and does not permit resistance (example from Sweetser (1988: 54), quoted in Hopper & Traugott (1993: 79)):

(6) I have to/??must get this paper in, but I guess I’ll go to the movies instead.

This explanation can be extended to epistemic *must* because this also expresses necessity (and hence involves strong commitment), although, as pointed out in section 4.2, *have to* is not a rival to epistemic *must*.

If there is a move away from forms which express a stronger commitment, then we would also expect other forms which express necessity to show a decrease. This is in fact the case: *should* and *ought to*, both classified by Coates (1983) as modals of necessity and obligation, show a decrease in frequency in DCPSE and in written corpora (see Leech (2003) and Leech *et al.* forthcoming for results from the Brown family of corpora).

The data from DCPSE cannot tell us which occurred first, the fall of root *must* or the rise of root *have to*. The figures for epistemic *must* suggest the former, otherwise it is difficult to explain why epistemic *must* has declined but epistemic *have (got) to* has not increased.

5. Conclusion

This study has investigated the variation between the core modal *must* and the semi-modals *have to* and *have got to* using the *Diachronic Corpus of Present-Day Spoken English*. By considering only syntactic environments in which the forms are variable, any possible competition between *must* and the “rival” semi-modal alternates can be reliably observed. The results from the study show that even in the 1960s data, the frequency of root *have to* has surpassed that of root *must*. The continued decline of

must and the rise in *have to* in the 1990s data suggests that *have to* is being used in contexts where previously we would have found *must*. Surprisingly, *have got to*, the most recent marker of root modality, has decreased. This does not appear to be a reflection of the DCPSE sampling, as Tagliamonte and Smith (2006) found that this was also the case in a number of English dialects.

The semantic coding carried out on *must*, *have to* and *have got to* in DCPSE has led to a number of interesting results. Firstly, although both epistemic and root senses of *must* have declined, as proportions of the total number of *must* they have remained fairly constant. This supports Leech's (2003) claim that *must* is not becoming monosemous. Study of the semantic senses of other modals in DCPSE is necessary in order to determine if this fact is peculiar to *must*. Secondly, *have to* is the most frequent marker of root modality, which is remarkable because we might expect the newer form *have got to* to be more frequent. And finally, epistemic *must* has decreased, but epistemic *have to* has not increased. This highlights the importance of using corpora to investigate language change, as simple 'anecdotal observation' would most likely have led to predictions that these forms are on the increase.

References

- Aarts, Bas, Gerald Nelson & Sean Wallis. 1998. "Using Fuzzy Tree Fragments to explore English grammar". *English Today* 14: 3: 52-56.
- Bauer, Laurie. 2002. "Hitting a moving target: An interim report on language change across generations". *English Today* 18: 4: 55-59.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Depraetere, Ilse & An Verhulst. 2008. "Source of Modality: A Reassessment". *English Language and Linguistics* 12: 1: 1-25.
- Fowler, Henry Watson. 1965. *A dictionary of modern English usage*. Second edition. Revised by Ernest Gowers. Oxford: Clarendon Press.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

- Hundt, Marianne. 1998. "It is important that this study (should) be based on the analysis of parallel corpora: On the use of the mandative subjunctive in four major varieties of English." *The Major Varieties of English* (Papers from MAVEN 97) ed. by Lindquist, Hans, Staffan Klintborg, Magnus Levin and Maria Estling, 159-175. Växjö: Acta Wexionensia,
- Kroch, Anthony S. 1989. "Reflexes of grammar in patterns of language change". *Language Variation and Change* 1: 199-244.
- 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 Lindquist, Hans, Staffan Klintborg, Magnus Levin & Maria Estling (eds.) *The Major Varieties of English* (Papers from MAVEN 97) ed. by Lindquist, Hans, Staffan Klintborg, Magnus Levin and Maria Estling, 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." *Modality in Contemporary English* ed. by Facchinetti, Roberta, Manfred Krug & Frank Palmer, 223-240. Berlin: Mouton de Gruyter.
- Leech, Geoffrey, & Jennifer Coates. 1980. "Semantic indeterminacy and the modals." *Studies in English linguistics* ed. By Greenbaum, Sidney, Geoffrey Leech & Jan Svartvik, 79-90. London: Longman.
- Leech, Geoffrey, Christian Mair, Marianne Hundt, & Nicholas Smith (forthcoming) *Change in contemporary English*. Cambridge: Cambridge University Press.
- Lightfoot, David W. 1979. *Principles of Diachronic Syntax*. Cambridge: Cambridge University Press.
- Mair, Christian. 1997. "Parallel corpora: a real-time approach to the study of language change in Progress." *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, ed. by Ljung, Magnus, 195-209. Amsterdam: Rodopi.
- Mair, Christian. 1998. "Corpora and the study of the major varieties of English: Issues and results." *The Major Varieties of English* (Papers from MAVEN 97)

- ed. by Lindquist, Hans, Staffan Klintborg, Magnus Levin and Maria Estling, 139-157. Växjö: Acta Wexionensia,
- Mair, Christian. 2008. "Corpora and the Study of Recent Change". *Corpus Linguistics: An International Handbook* ed. by Lüdeling, Anke & Merja Kytö. Berlin: Mouton de Gruyter.
- Mair, Christian & Geoffrey Leech. 2006. "Current changes in English syntax". *The Handbook of English Linguistics* ed. by Aarts, Bas & April McMahon, 318-342. Oxford: Blackwell.
- Myhill, John. 1995. "Change and continuity in the functions of the American English modals." *Linguistics* 33: 157-211.
- Nelson, Gerald, Sean Wallis & Bas Aarts. 2002. *Exploring natural language: working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.
- Övergaard, Gerd. 1995. *The Mandative Subjunctive in American and British English in the 20th Century* (Studia Anglistica Upsaliensia 94). Uppsala: Acta Universitatis Upsaliensis. Stockholm: Almqvist and Wiksell International.
- Pintzuk, Susan. 1989. "Variation and change in Old English clause structure." *Language Variation and Change* 7: 229-260.
- Smith, Nicholas. 2003. "Changes in modals and semi-modals of strong obligation and epistemic necessity in recent British English". *Modality in Contemporary English* ed. by Facchinetti, Roberta, Manfred Krug & Frank Palmer, 241-266. Berlin: Mouton de Gruyter.
- Svartvik, J. (ed). 1990. *The London Corpus of Spoken English: Description and Research*. Lund Studies in English 82. Lund University Press.
- Sweetser, Eve E. 1988. "Grammaticalization and semantic bleaching". *Berkeley Linguistics Society: General Session and Parasession on Grammaticalization*. ed. by Axmaker, Shelley, Annie Jaissner & Helen Singmaster.
- Tagliamonte, Sali & Jennifer Smith. 2006. "Layering, competition and a twist of fate: Deontic modality in dialects of English". *Diachronica* 23:2: 341-380.
- Traugott, Elizabeth Closs. 1989. "On the rise of epistemic meanings in English: an examples of subjectification in semantic change". *Language* 66: 1: 31-55.
- Wallis, Sean, and Gerald Nelson. 2000. "Exploiting fuzzy tree fragments in the investigation of parsed corpora". *Literary and linguistic computing* 15: 3: 339-361.

Warner, Anthony. 1993. *English Auxiliaries: Structure and history*. Cambridge: Cambridge University Press.

¹ For Quirk et al. (1985: 141-146) *have to* is a modal semi-auxiliary, while *have got to* is a modal idiom. These combinations are not identical in their syntactic behaviour: *have to* takes *do*-support, while inversion is not possible for most speakers. It can occur as a non-finite form and be preceded by an auxiliary verb. By contrast, *have got (to)* is always finite, and therefore cannot be preceded by auxiliary verbs. It can invert with a subject. We will use the label *semi-modal* for *have to* and *have got to*.

² The term 'Brown quartet' refers to four comparable corpora: the Brown Corpus and Frown (Freiburg-Brown) Corpus from 1961 and 1991-2, respectively, and two matching corpora of British English, LOB (Lancaster-Oslo/Bergen) and FLOB (Freiburg-Lancaster-Oslo/Bergen), again from 1961 and 1991-2, respectively. These corpora each contain a million words in fifteen written text types.

³ The SEU-mini corpus used by Leech (2003) contains 80,000 words (16 texts) from conversation, BBC discussions and news, sports and other commentaries, broadcast talks, etc., all dated between 1959-65. The ICE-GB-mini corpus contains 80,000 words from text categories chosen to match those of the SEU-mini corpus, the only difference being the date: texts are dated 1990-92.

⁴ Leech's (2003) results show that the epistemic reading of *may* has increased in the written corpora, but all other meanings have shown a decline. With *should* the root sense of "weak obligation" is shown to be gaining at the expense of other senses, particularly the epistemic sense of "weak inference."

⁵ See also www.ucl.ac.uk/english-usage/resources/ftfs for further information on conducting experiments using FTFs.

⁶ Depraetere and Verhulst (2008: 16) also exclude negative uses of *must* and *have to* based on the observation that "the scope of negation is different for *have to* (absence of necessity, 'not necessarily') and *must* (prohibition, 'necessarily not')."

⁷ There are two plausible reasons why these forms are not present in DCPSE: (i) they are very infrequent, or (ii) they are a post 1992 innovation. It is impossible to determine which, if either, of these is correct.

⁸ Example retrieved using WebCorp on 22/09/08. Web address: <http://www.webcorp.org.uk/cgi-bin/view.nm?url=http://www.unrealitytv.co.uk/x-factor/the-worst-x-factor-ever/&term=has%20to%20have%20been&case=case>.

⁹ Tagliamonte and Smith (2006) exclude these cases from their data based on the fact that they occurred exclusively with *must*, but in DCPSE *have got to* and *have to* can have performative meaning.

¹⁰ Coates's original mark-up has been removed from the example as it is not relevant here.

¹¹ This could be viewed as indirect support for the coding system chosen by Coates; we might expect that the better the coding system is, the fewer indeterminate examples there will be.

¹² The frequencies for epistemic *have to* and *have got to* are low which may lead to questions about the reliability of the chi-square test here, but log likelihood calculations confirm low significance scores: *must* 0.19, *have got to* 3.05, *have to* 1.67 (total 4.92).