

# **MUST and its rivals in the Diachronic Corpus of Present-Day Spoken English**

Jo Close (j.close@ucl.ac.uk) & Bas Aarts (b.aarts@ucl.ac.uk), UCL  
ISLE, Freiburg, 8<sup>th</sup> October 2008

## **1. Introduction**

In this paper we investigate recent change in the use of the semi-modals *HAVE TO* and *HAVE GOT TO* and the core modal *MUST* in a corpus of spoken English, with the aim of answering the following questions:

- Is there support for the idea that core modals are decreasing?
- If *MUST* is found to be decreasing, can the decrease be related to (an increase in) the use of “rival” semi-modal forms?
- Is there support for the theory that modals are becoming monosemous?
- What is driving the change in the use of the three forms?

We will compare the results from the spoken data with results from written corpora (LOB and F-LOB) presented in Leech et al. (forthcoming).

A further aim of this work is to contribute to the understanding of recent change, an aim we share with Mair, Hundt, Leech and Smith, the authors of the forthcoming CUP book *Change in contemporary English*.

## **2. The Diachronic Corpus of Present-Day Spoken English**

The *Diachronic Corpus of Present-Day Spoken English* (DCPSE) contains around 400,000 words from the British component of the *International Corpus of English* (ICE-GB) collected in the early 1990s and around 400,000 words from the *London-Lund Corpus* (LLC) collected between the late 1960s and early 1980s (<http://www.ucl.ac.uk/english-usage/projects/dcpse/>). Unlike the *FLOB* and *FROWN* corpora, compiled by Christian Mair at Freiburg, DCPSE contains only spoken English, because spoken English is where changes are likely to occur first.

“Spoken data [...] are the site of origin of almost all non-prestige innovations in language” (Mair, forthcoming 2008).

All the sentences in DCPSE have been grammatically analysed and have been given a detailed parse tree.

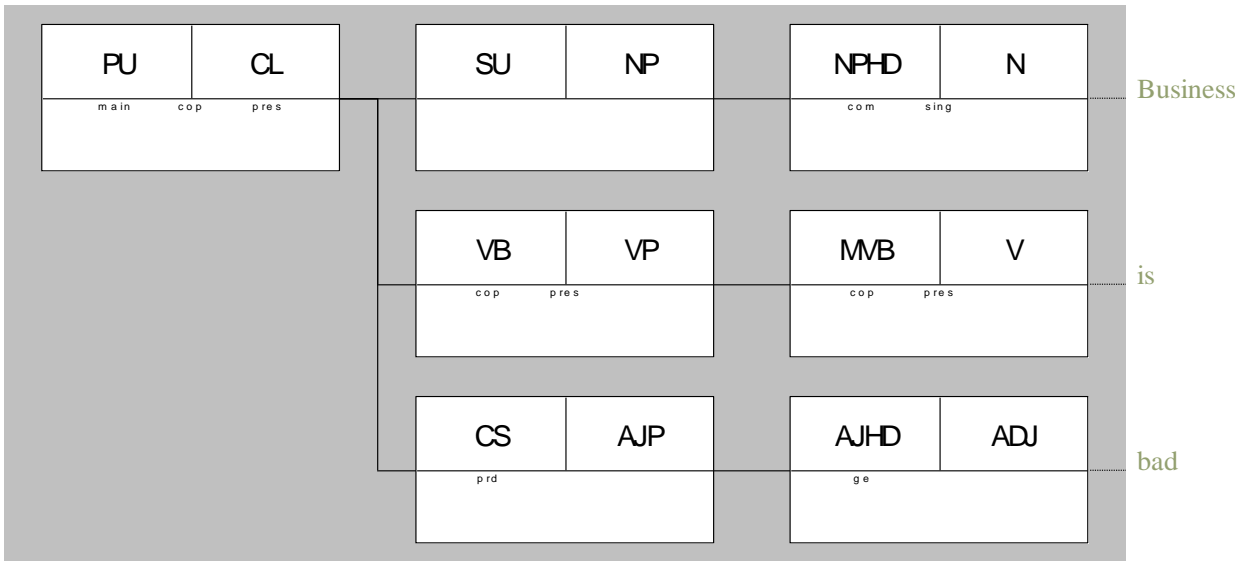


Figure 1. The grammatical analysis of the sentence *Business is bad* in DCPSE. PU= Parsing Unit, SU = Subject, VB = Verbal, MVB = Main verb, CS = Subject Complement, AJHD = Adjective Phrase Head, ADJ = Adjective.

Using the International Corpus of English Corpus Utility Program (ICECUP), it is possible to search DCPSE in a number of ways from simple text searches to more detailed grammatical queries using *Fuzzy Tree Fragments* (FTFs), which will retrieve matching examples from the corpus (Aarts, Nelson and Wallis 1998; Wallis and Nelson 2000; Nelson, Wallis and Aarts 2002). The FTF in figure 2 searches the corpus for the string *got to* with the position for *HAVE* left unspecified. We use the label *fuzzy* in Fuzzy Tree Fragments to indicate that users can be as precise or as vague as to what they wish to search for.

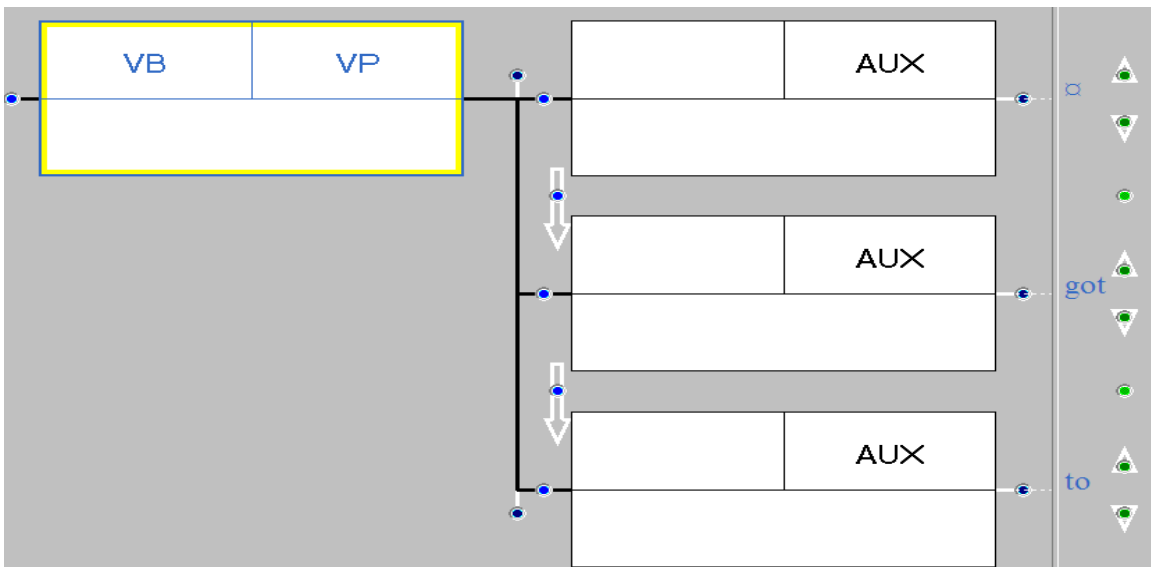


Figure 2. FTF for (HAVE) GOT TO.

### 3. Data and methodology

- Retrieval of data from corpus using text searches and FTFs.
- Data arranged by *source corpus* (i.e. date).
- Manual semantic coding (use of sound files to disambiguate).
- Statistical tests on data (chi-square, log likelihood).

Data excluded from the study in order to study only variable contexts:

- Negated forms (not semantically equivalent: *you mustn't go* "you are not allowed to go" vs. *you don't have to go* "you are allowed to not go").
- Interrogatives (only one example with *MUST*).
- Past tense forms of *HAVE TO* (no past tense of *MUST*).
- Non-finite forms *HAVE TO*, including future forms (no non-finite form of *MUST* or *HAVE GOT TO*).
- Unfinished/interrupted utterances.

### 4. Semantic coding

Manual semantic coding is necessary to test claims which refer to semantic meaning of the modals/semi-modals, such as the following:

- Leech (2003): *MUST* is one few modals not becoming monosemous.
- Coates (1983): *HAVE TO* and *HAVE GOT TO* are infrequent as epistemic modals.

Coding scheme of Leech and Coates (1980) and Coates (1983): two semantic classes of modals, Root and Epistemic, which have core and peripheral members.

#### *Epistemic MUST*

- *Logical necessity*: "In light of what is known, it is necessarily the case that x".
- *Confident inference*: "I confidently infer that x".

(Coates 1983: 41)

- (1) I am coming to conclude that it must be because I get too much water in it or something (DCPSE:DI-B47/ICE-GB:S1A-057 #0166:1:B).
- (2) Presumably one of his unsatisfied customers must have shopped him (DCPSE:DI-B58/ICE-GB:S1A-073 #0155:1:A).

#### *Root MUST*

- Basic meaning: "it is necessary for..."
- Strong obligation: "*It is imperative/obligatory that x*".
- Weak obligation: "*It is important that x*". (Coates 1983: 41)

- (3) My boss had said you must read those books (DCPSE:DI-B12/ICE-GB:S1A-016 #0171:1:E).
- (4) I must go and get a cardigan (DCPSE:DI-B63/ICE-GB:S1A-080 #0263:1:B).

*Performative/Speech-Act modals*

- (5) There 's a piece here called Spring Fire which takes my interest I must say (DCPSE:DI-D12/ICE-GB:S1B-032 #0141:1:A).
- (6) Uhm I must confess that I 'm unrepentant about the poll tax (DCPSE:DI-D14/ICE-GB:S1B-034 #0006:1:B).

*Ambiguity*

Coates (1983: 47): '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.'

- (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.  
 (= Root 'it's essential that mental health is related to ...'  
**Or** Epistemic 'it's inevitably the case that mental health is related to...')

**5. Results and discussion**

A comparison of the frequency of the three forms shows a significant decline in the use of *MUST* and a significant rise in the use of *HAVE TO*:

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	$\chi^2$ score
MUST	427	92.01	172	40.82	-55.64	<b>36.29</b>
HAVE GOT TO	187	40.30	156	37.02	-8.14	3.10
HAVE TO	188	40.51	225	53.40	+31.82	<b>31.94</b>
<b>TOTAL</b>	<b>802</b>	<b>172.82</b>	<b>553</b>	<b>131.24</b>	<b>-24.06</b>	<b>71.32</b>

Table 1. Overall frequencies of *MUST*, *HAVE GOT TO* and *HAVE TO* in DCPSE (figures in bold are significant at  $p < 0.01$ ).

Although both root and epistemic *MUST* show a decline, when viewed as proportions of total *MUST* (Table 2), root *MUST* remains constant (39.81% in LLC and 39.53% in ICE-GB), and epistemic *MUST* shows an

increase of just under 4% (from 47.78% in LLC to 51.74% in ICE-GB). There is little evidence that *MUST* is becoming monosemous.

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).

Leech et al. present the following distribution for *MUST* in written English (using the LOB and F-LOB corpora):

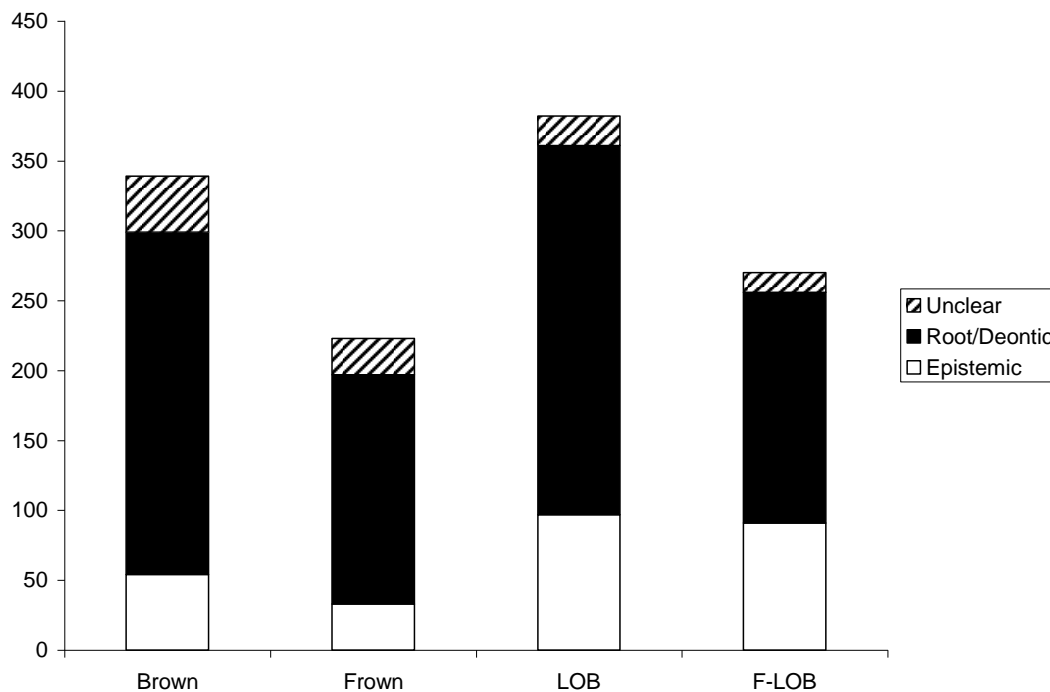


Figure 3. *Must* – change in frequency of senses (analysis of every third example) in the Brown family of corpora (Leech et al. forthcoming).

The general consensus in the literature is that the decline of the core modals cannot be attributed directly to an increase in the frequency of the semi-modals (Mair 1997, 1998; Krug 2000; Leech 2003; Smith 2003; Mair and Leech 2006). To discover if the decline in root and/or epistemic *MUST* is related to the use of the semi-modals *HAVE GOT TO* and *HAVE TO*, root and epistemic uses of the three forms were compared.

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	$\chi^2$ score
MUST	174	37.49	70	16.61	-55.69	<b>23.61</b>
HAVE GOT TO	184	39.65	151	35.84	-9.61	0.12
HAVE TO	185	39.86	208	49.36	+23.83	<b>12.32</b>
TOTAL	543	117	429	101.81	-12.98	<b>36.05</b>

Table 3. Frequencies of root *MUST*, *HAVE GOT TO* and *HAVE TO* in DCPSE (figures in bold are significant at  $p < 0.01$ ).

(Semi-) modal	LLC frequency		ICE-GB frequency		Change in frequency	
	raw	per 100,000 words	raw	per 100,000 words	%	$\chi^2$ score
MUST	206	44.39	91	21.60	-51.34	0.19
HAVE GOT TO	2	0.43	4	0.95	+120.93	3.36
HAVE TO	2	0.43	3	0.71	+65.12	1.83
TOTAL	210	45.25	98	23.26	-48.60	5.38

Table 4. Frequencies of epistemic *MUST*, *HAVE GOT TO* and *HAVE TO* in DCPSE.

The numbers of epistemic *HAVE TO* and *HAVE GOT TO* are low, but log likelihood calculations confirm low significance scores: *MUST* 0.19, *HAVE GOT TO* 3.05, *HAVE TO* 1.67 (total 4.92).

#### *What is driving the change?*

Smith (2003:259): "MUST is a casualty of a changing society where increasing emphasis is being placed on equality of power ...these conditions are likely to disfavour the use of MUST, they should correspondingly favour other forms which express obligation less directly."

Myhill (1995) suggests a growing tendency to avoid overt claims to authority by the speaker/writer, and claims this results in the decline in *MUST* (=obligation) and the rise of *SHOULD* (=weak obligation).

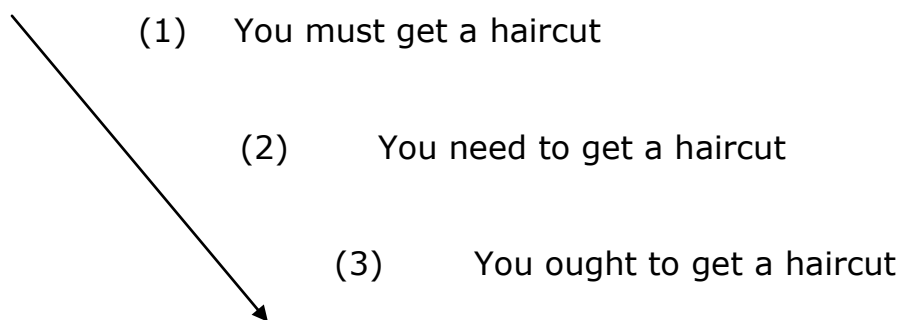
According to these statements made by Smith and Myhill, we might expect alternate expressions to increase in frequency.

- *NEED TO*: 3.66 LLC → 13.76 ICE-GB; +275.96%.
- *SHOULD*: 85.97 LLC → 72.38 ICE-GB; -15.81% (but shift towards root use, see table 5).

SHOULD	LLC frequency		ICE-GB frequency		Change in frequency %
	raw	per 100,000 words	raw	per 100,000 words	
EPISTEMIC	34	7.33	34	8.07	+10.1
ROOT	226	48.70	230	54.58	+12.07
OTHER	125	29.94	41	9.73	-67.5
TOTAL	385	85.97	305	72.38	-15.81

Table 5. Frequencies of *SHOULD* by semantic type in DCPSE ('other' includes formulaic expressions, cases where *should* is being used for *would* and ambiguous cases).

Leech (2004) suggests a scale of intensity for modal meaning:



*What is the future of the modal alternation?*

Krug (1998, 2000) argues for *GOTTA* as the main marker of deontic modality. There is no evidence for this in DCPSE where there are only 11 examples of *GOT TO* (2 in LLC; 9 in ICE-GB), perhaps only 8 of which are true examples (all in ICE-GB). Tagliamonte and Smith (2006: 373), however, show that some dialects of English are "holding on to" *HAVE TO*.

## 6. Conclusions

- There is a (significant) decline in the frequency of *MUST* as a marker of root obligation and epistemic necessity and a (significant) increase in the frequency of *HAVE TO* as a marker of root obligation in DCPSE.
- *MUST* is the dominant marker of epistemic necessity; epistemic cases of *HAVE GOT TO* and *HAVE TO* are very rare in DCPSE.
- *HAVE GOT TO* has declined in frequency in DCPSE.
- There is no change found with *SHOULD* which might account for the decline in frequency of *MUST*.

## References

- Aarts, Bas, Gerry Nelson and Sean Wallis (1998) Using Fuzzy Tree Fragments to explore English grammar. *English Today* 14, 52- 56.
- Coates, Jennifer (1983) *The semantics of the modal auxiliaries*. London: Croom Helm.
- 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*, 177–191. Växjö: Växjö University.
- Krug, Manfred G (2000) *Emerging English modals: A corpus-based study of grammaticalization*. Berlin/New York: Mouton de Gruyter.
- Leech, Geoffrey, and Jennifer Coates (1980) Semantic indeterminacy and the modals. In Greenbaum, Sidney, Geoffrey Leech and Jan Svartvik (eds.) *Studies in English linguistics*, 79-90. London: Longman.
- Leech, Geoffrey (2003) Modality on the move: the English modal auxiliaries 1961-1992. In Facchinetti, Roberta, Manfred Krug and Frank Palmer (eds.) *Modality in contemporary English*, 223-224. Berlin: Mouton de Gruyter.
- Leech, Geoffrey (2004) *Meaning and the English Verb*. Third edition. Harlow, England : Pearson/Longman.
- Leech, Geoffrey, Christian Mair, Marianne Hundt and Nicholas Smith (forthcoming) *Change in Contemporary English*. Cambridge: Cambridge University Press.
- Mair, Christian, and Geoffrey Leech (2006) Current changes in English syntax. In Aarts, Bas, & April McMahon (eds.) *The handbook of English linguistics*, 318-342. Oxford: Blackwell.
- Mair, Christian (forthcoming 2008) "Corpora and the study of recent change." In Lüdeling, Anke, Merja Kytö & Tony McEnery (eds.) *Corpus linguistics: An international handbook*. Berlin: Mouton de Gruyter.
- Myhill, John (1995) Change and continuity in the functions of the American English modals. *Linguistics* 33, 157-211.
- 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.
- Smith, Nicholas (2003) Changes in modals and semi-modals of strong obligation and epistemic necessity in recent British English. In Facchinetti, Roberta, Manfred Krug and Frank Palmer (eds.) *Modality in contemporary English*, 241-266. Berlin: Mouton de Gruyter.
- Tagliamonte, Sali, and Jennifer Smith. (2006) Layering, competition and a twist of fate: Deontic modality in dialects of English. *Diachronica* 23, 2, 341-380.
- Wallis, Sean, and Gerald Nelson (2000) Exploiting fuzzy tree fragments in the investigation of parsed corpora. *Literary and linguistic computing* 15, 3, 339-361.