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, 8th 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.
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.

![Diagram of grammatical analysis of "Business is bad" in DCPSE. PU=Parsing Unit, SU=Subject, VB=Verbal, MVB=Main verb, CS=Subject Complement, AJHD=Adjective Phrase Head, ADJ=Adjective.]

**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.

![Diagram of FTF for *(HAVE) GOT TO*.]

**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 monoseomous.
- 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:

<table>
<thead>
<tr>
<th>(Semi-)modal</th>
<th>LLC frequency raw per 100,000 words</th>
<th>ICE-GB frequency raw per 100,000 words</th>
<th>Change in frequency %</th>
<th>χ² score</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST</td>
<td>427</td>
<td>172</td>
<td>-55.64</td>
<td>36.29</td>
</tr>
<tr>
<td>HAVE GOT TO</td>
<td>187</td>
<td>156</td>
<td>-8.14</td>
<td>3.10</td>
</tr>
<tr>
<td>HAVE TO</td>
<td>188</td>
<td>225</td>
<td>+31.82</td>
<td>31.94</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>802</strong></td>
<td><strong>553</strong></td>
<td><strong>-24.06</strong></td>
<td><strong>71.32</strong></td>
</tr>
</tbody>
</table>

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 monoseamous.

<table>
<thead>
<tr>
<th>Source corpus</th>
<th>Epistemic</th>
<th></th>
<th>Root</th>
<th></th>
<th>Performative</th>
<th></th>
<th>Ambiguous</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLC</td>
<td>43.96</td>
<td>47.78</td>
<td>36.63</td>
<td>39.81</td>
<td>9.48</td>
<td>10.30</td>
<td>1.94</td>
<td>2.11</td>
<td>92.01</td>
</tr>
<tr>
<td>ICE-GB</td>
<td>21.12</td>
<td>51.74</td>
<td>16.14</td>
<td>39.53</td>
<td>2.37</td>
<td>5.81</td>
<td>1.19</td>
<td>2.91</td>
<td>40.82</td>
</tr>
<tr>
<td>TOTAL</td>
<td>65.08</td>
<td>48.99</td>
<td>52.77</td>
<td>39.73</td>
<td>11.85</td>
<td>8.92</td>
<td>3.13</td>
<td>2.35</td>
<td>132.83</td>
</tr>
</tbody>
</table>

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):

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 raw per 100,000 %  $\chi^2$ score
words words

<table>
<thead>
<tr>
<th></th>
<th>LLC frequency</th>
<th>ICE-GB frequency</th>
<th>Change in frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST</td>
<td>174</td>
<td>37.49</td>
<td>70</td>
</tr>
<tr>
<td>HAVE GOT TO</td>
<td>184</td>
<td>39.65</td>
<td>151</td>
</tr>
<tr>
<td>HAVE TO</td>
<td>185</td>
<td>39.86</td>
<td>208</td>
</tr>
<tr>
<td>TOTAL</td>
<td>543</td>
<td>117</td>
<td>429</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>LLC frequency</th>
<th>ICE-GB frequency</th>
<th>Change in frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUST</td>
<td>206</td>
<td>44.39</td>
<td>91</td>
</tr>
<tr>
<td>HAVE GOT TO</td>
<td>2</td>
<td>0.43</td>
<td>4</td>
</tr>
<tr>
<td>HAVE TO</td>
<td>2</td>
<td>0.43</td>
<td>3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>210</td>
<td>45.25</td>
<td>98</td>
</tr>
</tbody>
</table>

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).
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).

<table>
<thead>
<tr>
<th></th>
<th>LLC frequency per 100,000 words</th>
<th>ICE-GB frequency per 100,000 words</th>
<th>Change in frequency %</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPISTEMIC</td>
<td>34</td>
<td>7.33</td>
<td>+10.1</td>
</tr>
<tr>
<td>ROOT</td>
<td>226</td>
<td>48.70</td>
<td>+12.07</td>
</tr>
<tr>
<td>OTHER</td>
<td>125</td>
<td>29.94</td>
<td>-67.5</td>
</tr>
<tr>
<td>TOTAL</td>
<td>385</td>
<td>85.97</td>
<td>-15.81</td>
</tr>
</tbody>
</table>

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

(1) You must get a haircut

(2) You need to get a haircut

(3) You ought to get a haircut

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


