Ongoing grammatical change in spoken British English: real-time studies based on the DCPSE

Barbara Klein

1. Introduction

In the early 1990s, the completion of the Freiburg updates of the LOB and Brown corpora inaugurated a small boom in the corpus-based study of ongoing grammatical change in English in real time (see Mair 2006b: 12–35 for a discussion of methodological issues and Mair and Leech 2006 for a survey of important results). Given the composition of the “Brown family” of corpora, this research was necessarily restricted to the study of written English, and the extent to which diachronic developments observed in the written data reflected parallel trends in the spoken language remained controversial. The year 2006 saw the release of the *Diachronic Corpus of Present-Day Spoken English* (DCPSE), compiled by the Survey of English Usage at the University College London. The DCPSE is an 800,000 word corpus containing matching samples of spontaneous spoken British English from the London Lund Corpus (1958–77) and from the British component of the International Corpus of English (1990–92). For the first time, it is now possible to carry out “real time” studies on spoken data in roughly the same time span which is covered by the written Brown, LOB, Frown and F-LOB corpora.¹

The present paper uses the DCPSE to investigate three grammatical variables which have already been shown to be undergoing change in the written language: the s-genitive, the subjunctive, and a selection of modal

¹ For further information on the structure and availability of the Brown family of corpora see the homepage of ICAME, the International Computer Archive of Modern/Medieval English, at http://helmer.aksis.uib.no/icame.html. For information on additional corpus sources mentioned see David Lee’s “Bookmarks for Corpus-Based Linguists” at http://devoted.to/corpora. The corresponding information on the DCPSE can be obtained from http://www.ucl.ac.uk/english-usage/projects/dcpserecherche.htm.
verbs. In order to facilitate comparability to previous work based on the Brown family of corpora, I selected the late 1950s (1958–1960) and the early 1990s (1990–1992). Since thirty years are too short to track a grammatical change running its course from its primary emergence to its completion, this paper should be considered as the study of recent episodes in ongoing processes of change. The overall aim is to shed light on the extent to which changes proceed along parallel lines in speech and writing.

2. The s-genitive

Popular surveys on language change such as Barber (1964) or Potter (1969/75) have long argued that the frequency of the s-genitive is increasing in present-day English. Barber even claims a spread of the s-genitive to inanimate nouns, at the expense of semantically equivalent of-constructions (1964: 132). Even though these early studies made people aware of the issue, they are based on impressionistic observation only. Subsequent corpus-based research, however, has made clear that these impressions are not totally without foundation. Raab-Fischer (1995) investigates the press section of the LOB/F-LOB and the Brown/Frown corpora and shows that in newspaper language, the s-genitive is increasing with all kinds of possessor nouns, however not necessarily with those nouns ranking highest on the gender scale. The of-construction is shown to be losing ground with all types of possessor nouns. On the basis of an extensive elicitation study, Rosenbach claims that the s-genitive is increasing in present-day English, and that this extension is more productive than has been assumed so far (2002: 161). Mair (2006a: 245–246), too, confirms the general rise in the frequency of s-genitives in press and academic writing. However, he emphasizes that in his data the observed increase is not due to an increasing use of the s-genitive with inanimate nouns (which had been suggested as the responsible factor in other studies), but rather to a text-type specific desire in written genres to compress information. Given the relatively short period of investigation (c. 30 years), the statistical shifts are impressive. Mair and Leech (2006) as well as Leech and Smith (2006) report an increase in the frequency of s-genitives of 24.1%. Genitive-equivalent of-phrases, by contrast, decline by 23.4% (Mair and Leech 2006: 334). According to Leech and Smith, in present-day written British English, “the loss of of-genitives is very roughly commensurate with the gain of s-genitives” (2006: 197). Like

---

2 Due to the F-LOB corpus still being under construction at that time, only F-LOBpress was available for her research. Therefore, the analyses in the other corpora were restricted to the press section to allow full comparability of data.

3 The relative frequency of nouns from LOB to F-LOB remained relatively stable; Mair et al. (2002) documented only a slight increase by 0.8% in the relevant text types.
Raab-Fischer, Hinrichs and Szmrecsanyi (2007) base their work on the press section of the LOB/F-LOB and Brown/Frown corpora. Another paper (forthcoming$^4$) includes two non-matching spoken corpora (FRED and the CSAE$^5$). The authors argue that the increase in the s-genitive in press texts “may well reduce, at least partly, to an increasingly powerful tendency to code thematic NPs with the s-genitive” (2007: 468). They explain the increase in the s-genitive in terms of “economization”, the use of a more compact coding option in writing (2007: 469), as was also suspected by Mair (2006a: 245).

The question is whether grammatical change is documented at such massive rates in writing, are these changes also happening in speech? On the basis of the DCPSE, as I will show below, this question has to be answered in the negative.

For the identification of semantically equivalent s-genitives and of-constructions$^6$, I used the criteria listed in Hinrichs and Szmreczanyi (2007: 446f.), Kreyer (2003: 170) and Raab-Fischer (1995: 127). Table 1 presents the results and shows no discernible diachronic trend:

<table>
<thead>
<tr>
<th>DCPSE</th>
<th>1958–60</th>
<th>1990</th>
<th>Log likhd$^8$</th>
<th>Diff (%)$^9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total n/10,000</td>
<td>total n/10,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>s-genitive</td>
<td>67 18.00</td>
<td>74 18.13</td>
<td>0</td>
<td>+ 0.72</td>
</tr>
<tr>
<td>of-construction</td>
<td>88 23.64</td>
<td>78 19.11</td>
<td>1.88</td>
<td>- 19.17</td>
</tr>
</tbody>
</table>

s-gen 1958–60 vs. 1990: not significant; of-construction 1958–60 vs. 1990: not significant

---

4 The paper has not been published yet, but it is accessible on Szmrecsanyi’s website http://omnibus.uni-freiburg.de/~szmrecsa/
5 FRED (The Freiburg Corpus of English Dialects) contains dialectal speech by older speakers of British English, mainly recorded between 1970 and 1990. The CSAE (The Corpus of Spoken American English) consists of about 166,000 words of informal American English (Hinrichs and Szmreczanyi, forthcoming).
6 S-genitives were retrieved through the grammatical tagging of the corpus. For the of-constructions a Fuzzy Tree Fragment had to be created manually.
7 I limited the comparison to the data of 1958 to 1960 and the year 1990 for two reasons: first, it would have been far too laborious to extract the interchangeable of-constructions from the whole of the ICE-corpus (1990–1992), since the Fuzzy Tree Fragment identified 4,510 potential s-genitive equivalents. Second, the number of words in LUND 1958–1960 (37,225) almost exactly matches the number of words in ICE 1990 (40,819), which makes it an ideal basis for comparisons.
8 Log likelihood is a measure of statistical significance: a value of 3.84 or more equates with chi-square values >0.05; a value of 6.63 or more equates with chi-square values >0.01.
9 The column headed “Diff (%)” gives the increase (+) or decrease (-) in occurrences as a percentage of the frequency in the 1958–60 data.
The low log likelihood values demonstrate that, contrary to expectations, there is no significant change in the use of the forms in spontaneous spoken British English within the thirty years considered. The frequency of s-genitives remains stable, showing none of the drastic increase observed in writing. The use of the of-construction is decreasing slightly, yet below the degree of statistical significance.

It has been argued that the s-genitive is spreading to inanimate possessor nouns in the written register (cf. Barber 1964: 132f.). Whatever the merit of this explanation may be for written data, it is irrelevant for the spoken language. A look at the types of s-genitive possessors in the DCPSE reveals that there are no instances of inanimate possessors in the years considered for this study.

In sum, my analysis has illustrated that there is no diachronic change in genitive usage in spoken English. On the other hand, the increase of s-genitives in written English is so well documented that it cannot be doubted. The explanation for this discrepancy is that we are most likely not dealing with a genuine grammatical change here (in the sense of a change in the underlying system of choices and options), but with a stylistic development or an evolution of text-type specific norms in some written genres.

Colloquialization, the drift toward more oral modes of expression in writing, has been named as an explanation for the spread of s-genitives in writing (cf. Leech and Smith 2006: 197). However, the increase in locative, temporal (Raab-Fischer 1995: 126) or abstract inanimate possessors (Dahl 1971: 170) can hardly be explained by colloquialization, since the DCPSE clearly indicates that these forms are not very frequent in speech. Rather, the increasing frequency of the s-genitive reflects a growing demand for economy in language, and an increase in the information-density of some written genres.

2. The subjunctive

Toward the beginning of the 20th century, most grammarians considered the English subjunctive a moribund form on the verge of disappearing from the language completely (cf. Jespersen 1905: 205, or Fowler 1926: 595). In the second half of the century, a comeback was noticed, especially in American English (cf. Hirtle 1964, Jacobsson 1975). Övergaard (1995) is the first systematic diachronic study to look at the developments of the subjunctive in the 20th century, using different text corpora. She discovers a significant increase in the mandative subjunctive from 1960 to 1990 in British English, which she believes is due to American influence. In contrast to Övergaard (1995), Hundt (1998) is not able to confirm such a drastic increase in subj-
junctives comparing LOB to F-LOB. Her overall results for corpora representing different varieties,\textsuperscript{10} however, suggest that the subjunctive is becoming less formal (1998: 167). More recent studies by Serpollet (2001), Mair and Leech (2006), and Leech and Smith (2006) are all exclusively based on the Brown family of corpora. Serpollet (2001) reports a definite increase in the mandative subjunctive in British English, accompanied by a decline in periphrastic should.\textsuperscript{11} Mair and Leech (2006: 328) speak more cautiously of a “modest revival” of the present subjunctive in British English alongside a continuing decline of the were-subjunctive.

Apart from Hundt’s (1998) analysis of a sample of the spoken part of the BNC, there are no corpus-based studies on the use of the subjunctive in spoken English, let alone diachronic accounts of recent developments. In how far the spoken language has been affected by the revival of the mandative subjunctive is thus an issue that still needs to be looked into.

The following analysis of the DCPSE covers both the mandative and the past subjunctive and will also examine the formulaic subjunctive and the subjunctive in clauses of condition and concession. In the analysis of the mandative subjunctive, modal-verb periphrases and indicative forms will also be considered. Since the present subjunctive is morphologically distinct from the indicative only in the third person singular, many ambiguous forms are expected.\textsuperscript{12}

Table 2 shows the total number of subjunctives and their periphrastic and indicative alternants in subordinate clauses governed by mandative expressions:

\textsuperscript{10} LOB, Brown, F-LOB, Frown, WCNZE = Wellington Corpus of Written New Zealand English (1980s), ACE = Australian Corpus of English (late 1980s), BNC = British National Corpus (spoken part)

\textsuperscript{11} Since her results are not statistically significant due to low numbers, the proclaimed trends still need to be checked on the basis of larger corpora.

\textsuperscript{12} The verb be is an exception, because its base form be is distinct to all singular and plural present indicative forms. Past subjunctive were is only non-ambiguous in the first and third person singular.
Since the periods considered do not represent exactly the same number of words, the frequency of hits was normalized per 10,000 word tokens.

Hundt (1998) reports a decline of the mandative use of *should* by 17.28%, whereas Serpollet (2001) arrives at 46.45%, both relying on LOB and F-LOB.

Table 2: Distribution of the mandative subjunctive and its alternants in two sub-samples of the DCSPE

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total n/10,000</td>
<td>total n/10,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjunctive</td>
<td>0</td>
<td>0.00</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td><em>should</em></td>
<td>5</td>
<td>1.34</td>
<td>11</td>
<td>0.26</td>
</tr>
<tr>
<td>Other modals</td>
<td>2</td>
<td>0.54</td>
<td>9</td>
<td>0.21</td>
</tr>
<tr>
<td>Indicative</td>
<td>1</td>
<td>0.27</td>
<td>45</td>
<td>1.07</td>
</tr>
<tr>
<td>Non-distinct forms</td>
<td>0</td>
<td>0.00</td>
<td>10</td>
<td>0.24</td>
</tr>
<tr>
<td>Total mandative</td>
<td>8</td>
<td>2.15</td>
<td>77</td>
<td>1.83</td>
</tr>
</tbody>
</table>

* *should* 1958–60 vs. 1990–92: significant at p > 0.01

The fact that there are effectively only two genuine subjunctive forms in the 458,587 words analysed is a clear indication of its insignificance in speech. The subjunctive may be “alive and kicking” (Serpollet 2001: 531) in written English; it is still moribund, though, in contemporary spoken British English. The DCPSE data furthermore reveals that the frequency of the periphrastic construction with the modal *should* is on the decrease by 80.56% (significant at p > 0.01) in spoken British English, too. As the literature does not agree on the scope of the decline between the written corpora LOB and F-LOB due to different counting procedures, it can only be said that the decline of periphrastic *should* seems to be more pronounced in spoken English than in writing. In the DCPSE, modals other than *should* are also on the decline; one can however not generalize from these data, since the results are not statistically significant.

The results for the indicative and non-distinct verb forms are very interesting within the scope of the DCPSE, though not significant enough statistically to be used as a basis from which any broad trends can be drawn. The frequency of indicatives replacing subjunctives in mandative construction indeed increases by 297.55% from 0.27 in the late 1950s to 1.07 hits per 10,000 word tokens in the early 1990s. The question is, however, whether the non-distinct forms in the DCPSE actually represent genuine subjunctives or indicatives. It is generally agreed that ambiguous forms should be added to the numerically more important group, which in this case obviously are the

---

13 Since the periods considered do not represent exactly the same number of words, the frequency of hits was normalized per 10,000 word tokens.

14 Hundt (1998) reports a decline of the mandative use of *should* by 17.28%, whereas Serpollet (2001) arrives at 46.45%, both relying on LOB and F-LOB.
indicatives. Consequently, the inclusion of the non-distinct forms yields an increase in the frequency of indicatives by 385.89%\textsuperscript{15} (significant at $p > 0.05$). This result is in line with the study on written British English by Hundt (1998), who also claims that the indicative is gaining ground.

The synchronically most striking finding is that the indicative is the most frequent substitute for the genuine subjunctive in spoken British English of the early 1990s. Periphrasis by modals is used in about one quarter of all cases, while the genuine subjunctive is distinctly the least frequent option in mandative clauses.

As expected, the formulaic subjunctive turned out to be a relatively rare form. Nevertheless, it occurred more often than the genuine subjunctive. Even though one cannot speak of a drastic decline due to low numbers, there is still a slight decrease in its usage: down from 6 instances in the late 1950s to 4 examples in the early 1990s.\textsuperscript{16} The formulaic expressions used remain the same, ranging from simple *be it* to *far be it, so be it* and *if need be*. It looks as if these kinds of expressions are generally quite rare in spontaneous speech – probably due to their archaic nature.

The DCPSE offers very few examples of the use of the subjunctive in clauses of condition and concession. Although there is not a single form in the late 1950s, five attestations can be found in the years 1990 to 1992. Three are triggered by the expressions *if*, *whether* and *even though*, and involve the verb *be*. The other two examples contain the verb *come* used in a conditional sense.

The frequency of *were*-subjunctives has not changed dramatically in spoken English within the thirty years considered either. Against the background of a continuing decline in writing (cf. Mair and Leech 2006: 329), it could be expected that *were* would be losing ground in speech as well. In fact, there is a modest decline (by 16.07%), which, however, involves very low overall numbers and is thus not statistically significant. Interestingly, *was* is on the decline in hypothetical clauses as well, dropping by 53.3% (significant at $p > 0.05$). Therefore, the assumption that the decline in past subjunctives is counterbalanced by an increase in indicative *was* turned out to be not true for spontaneous spoken language. From a synchronic point of view, it is worth mentioning that indicative *was* is still the most frequent option in hypothetical contexts. In the years 1990 to 1992, 79.57% of all conditional and concessive statements in the first and third person singular preferred indicative *was* to the formal *were*-subjunctive. Consequently, Johansson and Norheim’s argument that the “*were*-subjunctive is clearly the dominant choice in hypothetical-conditional clauses” (1988: 34) did not turn out to be true for spontaneous speech of the early 1990s.

\textsuperscript{15} Log likelihood of 4.30

\textsuperscript{16} normalized per 10,000 word tokens: 1958–60: 1.61; 1990–92: 0.09.
It should be noted, moreover, that the main use of the *were*-subjunctive is in the formula *as it were*, which is used almost twice as much as the genuine past subjunctive in the years 1990 to 1992. Most of the 35 uses of *as it were* concentrate on the two subcorpora B (informal face-to-face conversation) and I (assorted spontaneous), where the expression serves as a type of discourse marker. The use of this form in spoken American English has not been investigated so far. Johansson and Norheim (1988: 34) were, however, able to demonstrate that *as it were* is much more frequent in LOB than in the Brown corpus. Whether *as it were* is a Briticism in speech as well is a question which can only be answered by a synchronic comparison of British and American spoken corpora.

Summing up, then, and looking at the results from a slightly different angle by combining the numbers of *should* and all other modals as well as the indicatives and the non-distinct forms, these are the salient developments in the thirty-year period under review:

**Table 3**: Distribution of *should* + other modals and indicative + non-distinct forms in mandative clauses

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total</td>
<td>n/10,000</td>
<td>total</td>
<td>n/10,000</td>
</tr>
<tr>
<td>Subjunctive</td>
<td>0</td>
<td>0.00</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td><em>Should</em> + other modals</td>
<td>7</td>
<td>1.88</td>
<td>20</td>
<td>0.47</td>
</tr>
<tr>
<td>Indicative + non-distinct forms</td>
<td>1</td>
<td>0.27</td>
<td>55</td>
<td>1.31</td>
</tr>
</tbody>
</table>

*should* + other modals 1958–60 vs. 1990–92: significant at p>0.01; indicative + non-distinct forms 1958–60 vs. 1990–92: significant at p>0.05

While, in terms of the subjunctive, one cannot speak of any developments, since it is virtually not used, the indicative is gaining ground extensively in spoken mandative clauses. One could even assume that the moderate loss in *should* and other modals is in parts compensated by the increase in indicatives. These results raise the question why the subjunctive is “alive and kicking” in written British English, but hardly used in spontaneous speech.

First of all, as in the case of the genitive, style and formality seem to play a pivotal role. Previous studies have shown that the development in written British English is remarkable, because the revival of the formal subjunctive cannot be explained by a colloquialization of the written norm. If it is thus due to an Americanization of British writing, one should be aware of the fact
that such a process is apparently not at work in spoken British English. Why is the subjunctive in speech not subject to Americanization? Probably, as it is still, even in American writing, a formal way of expressing commands or propositions, which is not typical of spontaneous speech. Övergaard (1995: 89) attributes the increase in mandative subjunctives to the impact of American texts after World War II. The present DCPSE analysis shows that this influence has not affected spoken language. It looks as though British writers are embracing the subjunctive, because it is an optional stylistic device, which offers variation in writing whenever a more formal option is more appropriate.

Like spoken British English, spoken American English has not been subject to real-time studies due to the lack of comparable corpora. Even though the role of the subjunctive in written American English has been thoroughly investigated, the only available source for speculations on the use of the subjunctive in speech are elicitation studies such as Nichols (1987) or Algeo (1988). These studies suggest that the subjunctive in American speech is not as rare as in British English, however, since these studies are not empirical and merely based on questionnaires, a diachronic set of comparable spoken corpora of American English is definitely required to find out, whether the two varieties are diverging or converging.

Once again, we have seen that a diachronic development in writing does not necessarily entail a corresponding trend in speech.

4. **Modal auxiliaries** *must, have to, (have) got to, need (to), shall*

As several corpus-based studies have shown, modality and the modal verbs have been “on the move” (Leech 2003) in the recent past, at least in written English (Leech 2003, 2004b; Mair and Leech 2006; Leech and Smith 2006). Unlike the *s*-genitive with inanimate nouns or the subjunctive, modals are extensively used in spoken English, which is why Leech (2003, 2004b) carried out a pilot study on a small subsection of what was to become the DCPSE. Even though these “mini-corpora” were too small to retrieve any statistically significant trends, it still became apparent that changes in the use of the modal auxiliaries should be much more pronounced in spoken English than in writing. Table 4 summarizes Leech’s results for the verbs studied also in the present paper:
Barbara Klein

Table 4: Frequencies of modals and semi-modals in written and spoken BE (Leech 2004b: 66ff.)

<table>
<thead>
<tr>
<th>Modals:</th>
<th>Written British English</th>
<th>Spoken British English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOB → F-LOB (Diff %)</td>
<td>mini corpora SEU → ICE-GB (Diff %)</td>
</tr>
<tr>
<td>must</td>
<td>- 29.0*</td>
<td>- 60.7*</td>
</tr>
<tr>
<td>shall</td>
<td>- 43.7*</td>
<td>- 34.6</td>
</tr>
<tr>
<td>need¹⁸</td>
<td>- 43.6*</td>
<td>(non-existent)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semi-modals:</th>
<th>Written British English</th>
<th>Spoken British English</th>
</tr>
</thead>
<tbody>
<tr>
<td>(HAVE) got to, gotta</td>
<td>- 34.1</td>
<td>- 25.7</td>
</tr>
<tr>
<td>HAVE to</td>
<td>+ 9.0</td>
<td>+ 31.6</td>
</tr>
<tr>
<td>NEED to</td>
<td>+ 249.1*</td>
<td>+ 650.0*</td>
</tr>
</tbody>
</table>

* significant at p > 0.01. Note: CAPITALIZED forms represent all morphological variants.

Leech’s comparison of LOB and F-LOB reveals that all the central modals relevant to this study appear to be decreasing in their frequency in written British English. Among the group of the semi-modals considered, there is no clear overall picture. Whereas need to is increasing drastically, the increase in have to is less pronounced. (Have) got to, on the other hand, seems to be declining.

Mair (2006b: 105) synchronically analyzed the Longman Spoken American Corpus and the conversation component of ICE-GB. He found out that have (got) to is much more common than must in the present-day speech of both varieties. He however attributes this change only to some extent to the fact that have to has additional syntactically-motivated uses. Like Smith (2003: 255), Mair claims that it is in particular the non-syntactically motivated forms that are increasing. Moreover, he concludes that have to is replacing must in its deontic use, but not in epistemic contexts (Mair 2006b: 105).

According to the previous studies it seems as if the two registers, though at a different pace, are heading in the same direction. Even though being closely aligned in terms of the number of words, the provisional two mini corpora used are not as closely matched as the DCPSE regarding time-frame and transcription files. The DCPSE is thus expected to allow a more in-depth examination of ongoing changes.

¹⁷ SEU: Survey of English Usage spoken “Mini-Corpus”. 80,000 words from selected categories of the years 1959–65. ICE-GB: International Corpus of English spoken “Mini-Corpus”. 80,000 words from selected categories of the years 1990–92.

¹⁸ Including needn’t, the form is decreasing by 40.2% (Leech 2003: 228).
Unlike the genitive or the subjunctive, the semi-modals are characteristic of spoken language, which leads to the assumption that innovations in their use originate in spontaneous speech, and subsequently spread to writing. My paper will be limited to the modals and semi-modals that have been reported to be most innovative. Since need to is reported to increase dramatically, the analysis of must and have (got) to is necessary, because changes in need to are very likely to affect other modals expressing the same concepts of obligation and necessity.

Shall is remarkable because it has been declining for at least two hundred years, and it will be interesting to find out which uses still survive in spoken British English.

Table 5 presents the frequencies of the modals and semi-modals of obligation and necessity, including negated forms and contractions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total n/10,000</td>
<td>total n/10,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>must</td>
<td>38 10.21</td>
<td>195 4.63</td>
<td>16.61**</td>
<td>- 54.67</td>
</tr>
<tr>
<td>(HAVE) got to</td>
<td>24 6.45</td>
<td>185 4.39</td>
<td>2.84</td>
<td>- 31.90</td>
</tr>
<tr>
<td>HAVE to</td>
<td>34 9.31</td>
<td>555 13.17</td>
<td>4.97*</td>
<td>+ 44.21</td>
</tr>
<tr>
<td>need</td>
<td>0 0.00</td>
<td>1 0.02</td>
<td>0.17</td>
<td>(+ 100.00)</td>
</tr>
<tr>
<td>NEED to</td>
<td>1 0.27</td>
<td>116 2.75</td>
<td>13.15**</td>
<td>+ 924.80</td>
</tr>
<tr>
<td>Total</td>
<td>97 26.06</td>
<td>1052 24.97</td>
<td>0.16</td>
<td>- 4.19</td>
</tr>
</tbody>
</table>

*HAVE to 1958–60 vs.1990–92: significant at p>0.05; **must, need to 1958–60 vs.1990–92: significant at p>0.01. CAPITALIZED forms represent all morphological variants.

The DCPSE data clearly indicate that, in spite of the overall stability, individual frequencies of modals and semi-modals of obligation and necessity have changed considerably within the thirty years examined. In the scope of this reallocation, core modal must has suffered most, dropping by 54.67% from 10.21 to 4.63 occurrences per 10,000 words. This decrease is not only statistically significant, but also much more pronounced than the decline in must in written British English (- 29.0%).

Interestingly, (have) got to, though being considered as informal and characteristically British (cf. Huddleston and Pullum 2002: 113), is declining in the DCPSE almost to the same extent as in writing from LOB to F-LOB (- 34.1%). The proportion of examples in which have is omitted is negligibly
small in both corpora. In the late 1950s corpus, *have* is dropped in none of the examples, and even in the early 1990s, *have*-omission amounts only to 3.78% of all instances of *got to*. It can however be suspected that even though *have*-omission is still rare, the phonetic reduction of *got to* gotta is already taking place in British speech.\textsuperscript{19}

The DCPSE furthermore shows that, unlike *must* and (have) *got to*, *have to* is increasingly used in British English, rising by 44.21%. While the written corpora only list an increase by 9.0%, Leech’s (2004b: 69) analysis of the spoken mini corpora (increase by 31.6%) has already foreshadowed the growing importance of *have to*.

In the case of *need*, the trends for auxiliary and main verb uses in spoken English are highly divergent. Auxiliary *need* followed by a bare verb form does not occur in the DCPSE. The only instance of modal usage is the negative contraction *needn’t*, which however only occurs once in the year 1992. On the basis of the DCPSE, it is thus possible to argue that *need* as an auxiliary verb remains virtually unused in present day spoken British English. Main verb *need to*, however, is drastically increasing in frequency, mounting from 0.27 to 2.75 occurrences per 10,000 words, which represents an extraordinary increase by 924.80%. The rise of *need to*, highly significant at the 1% level, is distinctly the most pronounced development among the modals and semi-modals. It is remarkable that this form was only very rarely used in the late 1950s, and is now ranking similar to (have) *got to* or *must*.

Similar to core modal *must*, *shall* has declined considerably in its frequency in spoken British English. The drop by 59.80% from the late 1950s to the early 1990s is statistically significant at p > 0.01.

### Table 6: Frequency of modal auxiliary *shall* in two sub-samples of the DCPSE

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total</td>
<td>n/10,000</td>
<td>total</td>
<td>n/10,000</td>
</tr>
<tr>
<td><em>shall</em></td>
<td>20</td>
<td>5.37</td>
<td>91\textsuperscript{20}</td>
<td>2.16</td>
</tr>
</tbody>
</table>

* significant at p>0.01

In the field of the modals and semi-modals of obligation and necessity, a close look at the synchronic distribution is necessary for the sake of comparing the frequencies of the individual forms. Table 7 lists the verbs *must, have*...
to, have got to, and need to in their order of precedence in written (LOB/F-LOB) and spoken British English (DCPSE). The numbers represent the normalized frequencies per 10,000 words.

Table 7: Modals and semi-modals of obligation and necessity in their order of precedence in speech and writing

<table>
<thead>
<tr>
<th>Rank</th>
<th>LOB</th>
<th>F-LOB</th>
<th>DCPSE 1958–60</th>
<th>DCPSE 1990–92</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/10,000</td>
<td>n/10,000</td>
<td>n/10,000</td>
<td>n/10,000</td>
</tr>
<tr>
<td>1</td>
<td>must</td>
<td>11.41</td>
<td>HAVE to</td>
<td>8.17</td>
</tr>
<tr>
<td>2</td>
<td>HAVE to</td>
<td>7.53</td>
<td>must</td>
<td>8.07</td>
</tr>
<tr>
<td>3</td>
<td>NEED (HAVE)</td>
<td>4.11</td>
<td>to</td>
<td>1.96</td>
</tr>
<tr>
<td>4</td>
<td>NEED to</td>
<td>0.54</td>
<td>(HAVE)</td>
<td>0.27</td>
</tr>
<tr>
<td>Total</td>
<td>23.59</td>
<td>18.50</td>
<td></td>
<td>26.06</td>
</tr>
</tbody>
</table>

**Must** and **have to** have changed places in both written and spoken British English within a thirty-year window. Whether this reversal is due to a replacement of must by have to is a question which cannot be answered straightforwardly, as there is a complex interaction of semantic and syntactic factors. The synchronic comparison of F-LOB and the 1990s DCPSE part furthermore shows that have to, as was suspected by Leech (2004b: 69), is used more frequently in speech than it is in writing. In addition, (have) got to, even though declining, is still more common in the spoken language of the early 1990s than in writing.

Another interesting result of this ranking is the fact that need to, even though increasing dramatically in speech, is still the least frequently used option among the four verbs.

Unlike the two preceding case studies, developments in speech and writing are broadly comparable with this variable. What divides writing and speech is not the presence or absence of a development but merely the speed at which these reorganizations in the modal system are happening: spontaneous speech clearly is more innovative here, and spoken corpora should therefore be the first choice for studying changes in the modals. Of

---

21 The numbers for the written corpora are adapted from the results by Smith (2003: 248), who normalized the frequencies per 100,000 words.
course, the statistical survey needs to be continued by qualitative follow-up studies.

The first step in the semantic analysis is the classification of the examples found according to various meanings and functions of the modals. Table 8 presents such a differentiated analysis for must in the DCPSE.

**Table 8: Meanings of must**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency</td>
<td>proportion</td>
<td>frequency</td>
<td>proportion</td>
</tr>
<tr>
<td>must epistemic</td>
<td>15</td>
<td>39.47%</td>
<td>93</td>
<td>47.69%</td>
</tr>
<tr>
<td>must deontic</td>
<td>23</td>
<td>60.53%</td>
<td>102</td>
<td>52.31%</td>
</tr>
<tr>
<td>total</td>
<td>38</td>
<td>100.00%</td>
<td>195</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

*must epistemic 1958–60 vs. 1990–92: significant at p>0.05; **must deontic 1958–60 vs. 1990–92: significant at p>0.01

As these data reveal, must is losing ground both in its epistemic and in its deontic function. The decline in deontic uses is slightly more pronounced. Again, the situation in written British English is similar: Smith (2003: 257) compares LOB to F-LOB, and reports a decline in both meanings of must. As in speech, the drop of epistemic must in writing is less severe (epistemic: 13.2%; deontic: 33.5%).

In the case of (have) got to, the overall usage in speech is different. Apart from one single exception all examples of (have) got to in the two parts of the DCPSE considered are deontic, which means that the reported decline can only be affecting its deontic uses. Since the meaning of deontic (have) got to is closer to must than to have to, the decline makes sense.

Surprisingly, all DCPSE examples of have to are deontic as well. Consequently, Leech’s (2004a: 81) assumption that epistemic have to is “growing more common in British English” appears to apply to written English exclusively. Since it is generally agreed that need to is rarely able to express epistemic modality, it is not surprising that it is likewise exclusively used in its deontic sense in the DCPSE.

The semantic decoding of the modals and semi-modals strongly suggests that the decrease in (have) got to and in the deontic uses of must is

---

22 <DI-B44_161> Loose shirts over jeans has got to be a sort of temporary prejudice, hasn’t it

23 From LOB to F-LOB, Smith (2003: 257) documented a significant increase of epistemic have to by 225.0%. Log likelihood 10.0; significant at p < 0.01.
compensated by the respective increase in have to and need to. Since must and need to cannot be used interchangeably in all contexts, and since (have) got to is not the regular alternative in past contexts, I will limit my comparison to deontic must and deontic, non-syntactically motivated have to. The major aim is to discover whether spoken British English is moving away from expressions of authority toward more polite and less pushy modes of expression. In order to retrieve only relevant and comparable forms, first all epistemic uses were eliminated from the overall number. In a second step, all syntactically motivated uses\textsuperscript{24} of have to were extracted, in order to guarantee comparability to bare verb form must. Table 9 shows the frequencies of all interchangeable forms.

**Table 9**: Deontic must and deontic, non-syntactically motivated have to

<table>
<thead>
<tr>
<th>DCPSE</th>
<th>1958–1960</th>
<th>1990–1992</th>
<th>Log likhd</th>
<th>Diff (n/10,000)</th>
<th>Diff (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>deontic must</td>
<td>23</td>
<td>6.18</td>
<td>102</td>
<td>2.42</td>
<td>13.43**</td>
</tr>
<tr>
<td>non-syntactically motivated HAVE to</td>
<td>22</td>
<td>5.91</td>
<td>392</td>
<td>9.30</td>
<td>4.92*</td>
</tr>
</tbody>
</table>

*HAVE to 1958–60 vs.1990–92: significant at p>0.05; **must 1958–60 vs.1990–92: significant at p>0.01

Capitalized HAVE here represents the present tense forms have and has. According to these data, it looks as if the increase in non-syntactically motivated have to, which is by definition in the majority of cases interchangeable with must, compensates for the decrease in deontic must. As we have seen before, must is not threatened by have to in epistemic function in spoken English. In regular finite, non-negated present tense, it appears as if have to was becoming more and more popular in present-day British English speech. However, it is absolutely essential to point out that it would be premature to deduce that there is a one-to-one replacement of must by have to. It is obvious that need to is also replacing must in some deontic contexts in the same way in which (have) got to can also be used instead of have to.

\textsuperscript{24} According to Smith (2003: 254), these are non-finite forms, negated forms expressing absent requirement - as in you don’t have to - and all past forms.
The case of declining *shall* is different from the changes in the modals and semi-modals discussed so far, because over the centuries, some of its uses have gradually disappeared (cf. Gotti 2003: 278f.). Table 10 shows the development of the two main remaining usages in the second part of the twentieth century:

Table 10: Meanings of *shall*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency</td>
<td>proportion</td>
<td>frequency</td>
<td>proportion</td>
</tr>
<tr>
<td><em>shall</em> volition</td>
<td>4</td>
<td>20.00%</td>
<td>43</td>
<td>47.25%</td>
</tr>
<tr>
<td><em>shall</em> prediction</td>
<td>16</td>
<td>80.00%</td>
<td>48</td>
<td>52.75%</td>
</tr>
<tr>
<td>total</td>
<td>20</td>
<td>100.00%</td>
<td>91</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

*shall* prediction and total 1958–60 vs.1990–92: significant at p>0.01

According to the DCPSE data, *shall* has been losing ground both in expressing volition and prediction. The decline in its futurity meaning is clearly more distinct, which is probably due to modal *will* becoming more popular and replacing *shall* in future contexts. What is striking is the high number of “formulaic discourse structuring devices” (Mair 2006b: 102) of the type *as we shall see*, which are not productive, and thus hint at the fact that *shall* might become obsolete in the course of the following centuries.

Since they encompass both the spoken and the written language, the developments in the field of modality probably reflect more genuinely grammatical changes than the shifts in the use of the genitive and the subjunctive. With the “emerging semi-modals” (cf. the title of Krug 2000) *have to* and *need to*, we are seeing *grammaticalization* at work.

**Colloquialization** is “a tendency for features of conversational spoken language to infiltrate, and spread in the written language” (Leech 2004b: 75). Thereby, *have to* and *need to* provide prime examples. Their rate of increase in speech is three times that in writing, which suggests that the trend originated in spontaneous speech and subsequently spread to writing, where it is thus still lagging behind somewhat.

**Democratization**, a “tendency to avoid unequal and face-threatening modes of interaction” (Leech 2004b: 75) is a socio-psychological model of language change, which can be perfectly applied to the decline in *must* and the respective rise in *have to* and *need to*. Myhill (1995), investigating the functions of a selection of modals before and after the American Civil War,
argues that the "old" modals *must*, *should*, *may*, and *shall* were associated with hierarchical social relationships, with people controlling the actions of other people, and with absolute judgements based upon social decorum, principle and rules about societal expectations of certain types of people. The 'new' modals [*i.e. have to, got to*] are more personal, being used to give advice to an equal, make an emotional request, offer help or criticize one's interlocutor (Myhill 1995: 157).

Even though my paper is neither focused on American English nor on the time around the Civil War, Myhill's description of the modals perfectly fits the trends observable in the DCPSE in the second half of the twentieth century. It is highly likely that the decline in *must* is the reflection of a societal tendency arising after World War II (cf. Mair and Hundt 1997: 78), which is characterized by a movement away from exercising authority openly and encourages more polite and egalitarian ways of formulating commands, orders or requests.

I conclude that in this area, changes in speech can be definitely considered prior to writing. Unlike the genitive and the subjunctive, the modals and semi-modals turned out to be not a mere stylistic option in writing, but a category of verbs whose origin lies in what the DCPSE represents: spontaneous spoken language.

5. Conclusion

On the basis of the DCPSE corpus, this paper has identified and analyzed recent developments in the use of three grammatical variables within the span of one generation. They have turned out to be fairly heterogeneous both in the scope and in the potential causes of the observed changes.

The genitive represents an area in which there is no significant ongoing change in present-day spoken English. Contrary to expectations raised by studies of written English, the frequencies of both s-genitives and equivalent *of*-constructions have remained stable within the thirty years documented in the DCPSE. Similarly, the use of the subjunctive has turned out to be more stable in speech than in writing. In speech, the present subjunctive is close to non-existent in British English. In mandative statements, the indicative is gaining ground, with even the *should* option declining. With modals and semi-modals, on the other hand, the diachronic dynamic seems to be greater in speech, with the changes originating there eventually being taken up in writing. This paper therefore provides evidence that one should not act on the general assumption that changes in speech always precede changes in writing. It appears as though changes in formal or stylistically optional
grammatical elements in writing may progress independently of the spoken register.

Beyond the immediate focus on three grammatical variables, the present paper can also be seen as a test for a new corpus resource, the DCPSE. In the compilation of this corpus, considerable efforts have been made by the compilers to control all variables in aligning the “older” London-Lund components with the “newer” ICE-GB ones, which has generally been successful. But some shortcomings remain. One is the years covered in the two parts of the corpus: while the 400,000 “new” words from ICE-GB cover three years, the “old” LUND part is spread over nineteen years. Since the first priority of my study was an exact match of years in order to guarantee comparability to the written corpora, I had to work with corpus parts of different sizes. As a consequence, text categories could not be completely matched.

This study is one of the first to be carried out on the basis of the DCPSE, and there will be more to follow. The field of phonetics will gain from this new corpus as well, since auditory files of the original recordings are included. In the vast field of morphosyntax, this paper can be considered as a mere starting point. As we have seen, the nature of the changes in speech and writing can vary considerably, which calls for further comparisons of the two registers using the DCPSE. In conclusion, it is imperative to emphasize the need for comparable corpora of spoken American English to create a set of spoken corpora similar to LOB and Brown, and their respective updates. In many ways, an American counterpart to the DCPSE would be the last piece of the puzzle, completing and uniting the existing corpora, and serving as a unique basis for an integrated study of regional variability and diachronic change in the recent history of spoken English.

References


Ongoing grammatical change in spoken British English


Jespersen, Otto (1905). Growth and Structure of the English Language. Chicago: Chicago UP.


Barbara Klein
Englisches Seminar
Universität Freiburg