Growth and Decline: How Grammar Has Been Changing in Recent English

Geoffrey Leech
Lancaster University
g.leech@lancaster.ac.uk

Abstract

Grammatical change, as opposed to (for example) phonetic or lexical change, is generally felt to be a slow process, imperceptible to the native speaker. There is some truth in this judgement, and yet changes in English grammatical usage during the recent past have been considerable. The evidence for this comes from comparable corpora or subcorpora systematically sampled from the language at regular intervals of time. It has now become possible to

- observe changes in progress in written English
- compare rates of change from one genre to another
- compare rates of change across the two regional varieties: American English (AmE) and British English (BrE)

1. The Brown Family of Corpora

The corpora I shall refer to, in demonstrating this, are known collectively as the Brown family of corpora. The name derives from the Brown Corpus, the earliest electronic corpus of the English language, collected at Brown University 1962-4, and since then widely distributed among researchers, teachers and students. The Brown Corpus consists of just over a million words, and is subdivided into 500 written (printed) text samples of approximately 2000 words each.

Alongside the Brown Corpus, the Lancaster-Oslo-Bergen corpus was created as a British counterpart, to be as precisely comparable to the Brown Corpus as possible. It consists, like Brown, of 500 samples of written English, constructed according to the same corpus design, with subdivision into 15 genres, within which individual texts and groups of texts are matched. Like those of the American corpus, the text samples of the British corpus were selected by stratified randomized sampling from texts first published in 1961.

It will be noticed that these corpora, like others subsequently created, were named after the universities at which they were compiled. Also, for convenience,
corpora created at more than one university have been informally identified by acronyms, such as “LOB corpus” for the “Lancaster-Oslo-Bergen corpus”.

Considering the immense growth in computing power and text storage capacity since the 1960s and 1970s, the Brown and LOB corpora are nowadays regarded as somewhat small and as well as being limited to written English. Nevertheless, they are still valuable for research because of their precise comparability, and also because of their wide availability: researchers all around the world can obtain copies of these corpora, and can annotate them and analyse them for their own research purposes.

Table 1.
The genres of the Brown family of corpora, and the number of text samples each contains.

<table>
<thead>
<tr>
<th>Identifying letter</th>
<th>Genres</th>
<th>Subcorpora</th>
<th>Number of text samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Press: reportage</td>
<td>Press</td>
<td>44</td>
</tr>
<tr>
<td>B</td>
<td>Press: editorial</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>C</td>
<td>Press: reviews</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>D</td>
<td>Religion</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>E</td>
<td>Skills and hobbies</td>
<td>General Prose</td>
<td>36</td>
</tr>
<tr>
<td>F</td>
<td>Popular lore</td>
<td></td>
<td>48</td>
</tr>
<tr>
<td>G</td>
<td>Belles letters, biography, memoirs etc.</td>
<td></td>
<td>75</td>
</tr>
<tr>
<td>H</td>
<td>Miscellaneous [mostly government documents]</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>J</td>
<td>Learned [academic writing]</td>
<td>Learned</td>
<td>80</td>
</tr>
<tr>
<td>K</td>
<td>General fiction</td>
<td>Fiction</td>
<td>29</td>
</tr>
<tr>
<td>L</td>
<td>Mystery and detective fiction</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>M</td>
<td>Science fiction</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>N</td>
<td>Adventure and western fiction</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>P</td>
<td>Romance and love story</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>R</td>
<td>Humour</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>500</td>
</tr>
</tbody>
</table>

Because the genre sections are relatively small, it has been found that more reliable comparisons can be made by grouping the genres into four subcorpora (Press, General Prose, Learned and Fiction), containing a set of similar genres, as shown in the third column of Table 1.

Table 2 summarizes the contents of each corpus of the Brown family, referring to the subcorpora which are used in subsequent data analyses in this paper.
Table 2.
The composition (in terms of major subcorpora) of the Brown Corpus and other members of the Brown family of corpora.

<table>
<thead>
<tr>
<th>Press:</th>
<th>Genres A, B and C</th>
<th>c. 178,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Prose:</td>
<td>Genres D, E, F, G and H</td>
<td>c. 413,000 words</td>
</tr>
<tr>
<td>Learned/Academic:</td>
<td>Genre J</td>
<td>c. 159,000 words</td>
</tr>
<tr>
<td>Fiction:</td>
<td>Genres K, L, M, N, P and R</td>
<td>c. 257,000 words</td>
</tr>
<tr>
<td>Each whole corpus is just over 1,000,000 words</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since the Brown and LOB Corpora were compiled, the Brown family has grown by the addition of the Freiburg-Brown Corpus and the Freiburg-Lancaster-Oslo-Bergen Corpus (Frown and FLOB for short), which again precisely match Brown and LOB, but with data sampled from publications in 1992 (American English) and 1991 (British English) respectively. In effect, these four corpora provide a means of tracking changes in usage over the period roughly of one generation, between the 1960s and the 1990s. Further, for the purpose of investigating recent change in standard written English, further comparable corpora have since been added, as shown in Table 3:

Table 3.
The evidence of change: comparable corpora of American English (AmE) and British English (BrE) at roughly 30-year intervals.¹

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>British English</td>
<td>BLOB-1901 Corpus</td>
<td>BLOB-1931 Corpus</td>
<td>LOB Corpus</td>
<td>FLOB Corpus</td>
<td>BE06</td>
</tr>
<tr>
<td>American English</td>
<td>(no corpus yet)</td>
<td>B-Brown Corpus</td>
<td>Brown Corpus</td>
<td>Frown Corpus</td>
<td>AmE06</td>
</tr>
</tbody>
</table>

¹ In the short names of the corpora, the initial “B” stands for “before”, and the initial “F” stands for “Freiburg”. Although it is clearly impossible to give due acknowledgement to all those who participated in the development of these corpora, the following list names those who took a leading role in the compilation of the corpora bracketed after them: Nelson Francis and Henry Kučera (Brown); Geoffrey Leech and Stig Johansson (LOB); Christian Mair and Marianne Hundt (Frown, FLOB); Nick Smith, Paul Rayson and Geoffrey Leech (BLOB-1931, BLOB-1901); B-Brown (Marianne Hundt), Paul Baker (BE06), Paul Baker, Amanda Potts and Matteo Di Cristoforo (AmE06).
On this diagram, all-white cells represent comparable corpora that are complete, and match in detail the original Brown Corpus, except for their date and regional provenance. The two shaded corpora on the left are at present incomplete. The BLOB-1901 corpus has been used in some of my analyses – but, because only a comparable but incomplete 1/3 of the corpus was available, the frequency data derived from that corpus had to be scaled up to be equivalent to the figures for the other corpora. The two lightly-shaded corpora on the right were collected from matching textual materials on the World Wide Web, and although in every respect they match the Brown family design, their web origin might conceivably have an influence on the style of language used, and hence of their comparability with the other corpora. (The reduced width of the rightmost column of Table 2 is a reminder that the most recent period represented by the Brown family, the period 1991-2006±2, is approximately only half as long as the 30-year periods separating the earlier corpora.) Obviously the results obtained from the shaded corpora have to be treated with some caution. The dates “1931±3” etc. indicate that for the more recently-collected corpora it was considered impractical to compile the corpus from data published in a single year (as had originally been done for the Brown, LOB, Frown and FLOB corpora). Instead, the three years preceding and following the target year were accepted as a window of time from which the corpus material could be collected. Thus “1931±3” for the BLOB-1931 and B-Brown corpora indicates that the window within which the text samples could be collected was 1928-1934. It was found that the task of collecting exactly matching texts from a particular year is very difficult and time-consuming, particularly for the earlier periods of the Brown family, and it was decided that the change of grammatical usage within a period as short as seven years was likely to be fairly negligible.

2. How Frequency Data Reflect Change

To illustrate how data from the Brown family can trace changes in the language during the twentieth century, I take the straightforward example of the preposition upon, which has declined progressively and markedly in frequency between the 1901 and 2006 corpora in British English, so that today it is less than one-sixth as frequent as it was at the beginning of the twentieth century.

Figure 1 shows the overall trend, measured in the number of occurrences of the item concerned (here upon) per million words. For the time being, we will accept this per-million-word measure (abbreviated to ‘pmw’) as the most convenient, though not necessarily the best, method of calculation. Figure 2 looks in more detail at the period 1931-1991, dividing the frequencies of the whole corpora into frequencies pmw for the four subcorpora Press, General Prose, Learned and Fiction.
The main extra information we obtain from Figure 2 is that (a) the decline of *upon* is general across the four subcorpora, but (b) that it is steeper, especially in the period 1931-1961, in certain subcorpora than in others. Thus the steepest decline is found in the Press subcorpus, which is not surprising, given that journalistic writing is often in the vanguard of change. On the other hand, the Learned variety, representing the conservative style of academic writing, shows...
the least change, so that by 1991 this subcorpus has ‘lagged behind’ and retains the highest frequency of upon.²

We may compare two methods of tracking change: the normalized frequency (pmw) method already illustrated, and the proportional (variable-and-variants) method which has been considered linguistically preferable, although it is often more problematic (see Smith and Leech, forthcoming). To illustrate the two methods, we show another instance of change, this time of marked increase. Figures 3 and 4 show the same phenomenon: the increasing use of negative contractions (–n’t instead of not) and contractions of auxiliary verb or copula. Figure 3 simply shows the increase use of -n’t in terms of instances per million words (pmw):

As in Figure 2, the chart shows a general increase across the four subcorpora, except that the Learned variety (again, more conservative) shows a scarcely perceptible rise from zero, even declining slightly but non-significantly towards zero in the last 15-year period. It is also not surprising that the highest frequency and the steepest rise are in the Fictional subcorpus, which in general shows the closest similarity to spoken English, of which contractions are most characteristic.³

---

² Hundt and Mair (1999) have described the contrasting slower and faster tendencies of ‘agile’ and ‘uptight’ genres (e.g. Press and Learned subcorpora respectively) in tracking linguistic change in corpora.

³ On increasing use of contractions in newspaper texts, see Axelsson (1998).
Figure 4 shows the increase in contractions employing the proportional method of calculation. This means that frequency of the item concerned is measured against what would be the frequency of that item, if all negatives after an auxiliary were contractions. The frequency of negative contractions is here registered as a percentage of potential contractions. To show this, we sum occurrences not only of didn’t, isn’t, hadn’t, etc. but also of did not, is not, had not, etc.

![Proportional method: Increasing frequency of negative contractions as a percentage of potential contractions, 1931-2006.](image)

It will be seen that the two methods of tracking change show very similar results. However, in the rest of the paper, I will find it convenient, and sometimes necessary, to employ the normalized frequency method (pmw), since the alternative proportional method is often problematic, in that one has the problem of (a) determining, and (b) calculating, the number of potential, as well as actual occurrences. This is relatively straightforward in the case of negative contractions, in that one has to search the corpus for V + not as well as V+n’t, where V is a contractable verb. Even here, however, difficulties arise: for example, am not and may not are not able to contract to amn’t and mayn’t in contemporary English, although they belong to the class of verbs (copula and modals) which are normally contractable. Without going into detail into the reasons why the alternation -n’t - not does not apply to all cases, it will be clear that ‘contractability’ is not an easy property to define, let alone to quantify, even though it might be thought one of the simplest grammatical variables to handle. In other cases, it is virtually impossible to define the variable. For example, the modal auxiliary must has been declining rapidly in recent English, but to define the class of expressions of ‘obligation/necessity’ which are variants of the same
variable as *must*, in the sense that they could replace *must* without significant change of meaning, is extremely difficult. In different contexts it could be *have to, need to, should, be obliged to, be necessary that*, etc., but these have different syntactic privileges of occurrence and vary somewhat in meaning. The number of competing variants is virtually open-ended.

3. Possible Causes of Frequency Change

The question as to why such frequency changes take place is another issue which it is difficult to resolve. Are the observed changes due to internal, linguistic reasons, or to external social reasons? Are they brought about by conscious human influence, or by unconscious means? Although these questions are impossible to answer definitively, we can put forward conjectural explanations for change by noting related tendencies of growth or decline taking place across a range of different grammatical phenomena. The chief processes we postulate are:

a. **Colloquialization (or informalization)** – adapting written language towards spoken norms (e.g. the increasing use of contractions, including negative contractions as already discussed)
b. **Grammaticalization** – the process of converting lexical into grammatical material (e.g. historically, the growth in the use of the progressive and of semi-modals such as *have to*)
c. **Americanization** – following the American lead – adopting American linguistic habits or conventions (e.g. the revival of the mandative subjunctive in British English)
d. **Densification** – packing more meaning into less space – as found particularly in press headlines
e. **Prescriptivism** – language change under the influence of overtly expressed attitudes either in favour of or against particular usages (e.g. the decrease of the use of the passive in recent decades).

Of these, *a, c, d* seem to be processes driven by social factors outside language, whereas grammaticalization (*b*) is chiefly internal to language (although spurred on by external influences). Prescriptivism (*e*) is the only case where conscious human motivation is involved. We will now consider each of these briefly in relation to the changing phenomena for which they offer an explanation.

3.1. Colloquialization

The trend whereby written language becomes more colloquial or speech-like has already been illustrated in the increasing use of negative contractions. Other
changes which appear to show the influence of colloquialization are increasing use of verb contractions, of semi-modals (see 3.2 below) of the present progressive (again see 3.2), of not-negation rather than no-negation (e.g. *She didn't say anything* rather than *She said nothing*). The shift of preference away from *wh*-relativization towards *that*- and zero relativization can also be partially attributed to colloquialization. Indeed, colloquialization, although not a general trend within the language, may be considered a ‘natural’ process, in the sense that innovations in the language frequently occur first in the spoken medium and then to spread to the written medium. As the instance of relative clauses shows, the growth of one form (the *that* relativizer) may be associated with the decline of another, competing with it (the *wh*-relativizer). This negative side of colloquialization can arguably be a partial explanation for the declining use of the passive (see 3.5 below).

3.2. Grammaticalization

The process of grammaticalization has been extensively investigated as one of the major factors accounting for language change, often over a long time-scale (see Hopper and Traugott 2003). Thus two textbook examples of grammaticalization come from the modal and aspectual systems of the English verb: the English modal auxiliaries (a rather well-defined class of verbs morphologically, syntactically and semantically) arose in a process of grammaticalization which reached its crucial period in late Middle English. More recently, a new ‘wave’ of grammaticalization has brought to prominence so-called semi-modals (or emerging modals – see Mair 1997, Krug 2000) such as *be going to*, *have to*, and *want to*. Grammaticalization is associated with increase in frequency, which can continue long after the main grammaticalizing process in other respects has concluded. These three semi-modals, with the addition of *need to*, have indeed been increasing steeply during the twentieth century, and one argument is that these have been encroaching on the more ‘traditional’ modals such as *must* and *will* (see Leech 2003, Smith 2003). The second case of grammaticalization we consider is the progressive aspect, which again appears to have arisen through grammaticalization in late ME and early ModE. Since then the progressive has gradually established itself in the language, and continues to increase in frequency, as is shown especially by the present progressive, which has increased c. 30% in both AmE and BrE over recent decades (Leech et al. 2009).

It seems likely that the processes of grammaticalization and colloquialization act in synergy or in tandem – for example, the growing use of semi-modals is far more evident in the spoken language – where especially in American
conversation *have to*, *got to* and *need to* are all nowadays more common than *must*, the modal auxiliary in the same semantic field of obligation/necessity. Grammaticalization involves phonetic processes (e.g. vowel reduction, ellipsis and assimilation, as in the reduced forms of semi-modals informally spelt *gonna*, *gotta*, and *wanna*) associated with the spoken language, and it is therefore not unreasonable to suppose that changes through grammaticalization arise primarily in the spoken language and gradually affect the written language through colloquialization.

### 3.3. Americanization

The influence of American English on other regional varieties, such as British English, is easily accepted on the level of lexis – where in this author’s lifetime American usages such as *movie*, *guys* and *I guess* have been steadily taking over from corresponding British usages *film*, *chaps/blokes*, and *I reckon/suppose*. To explain the largely one-way traffic from AmE to BrE, we hardly need to mention that the USA is easily the most influential English-speaking nation economically, culturally, and politically. Perhaps more relevant is the fact that the USA has by far the largest native English-speaking population – several times larger than the English-speaking population of the UK, the next most populous nation of the English-speaking world.

Americanization is therefore a familiar process to speakers of BrE, as of other varieties of English around the world, even if the word itself is rarely used. The question to ask here is: can Americanization be applied not only to lexical change but to grammatical change? Change in grammar, more than change in lexis, takes place at a level below conscious notice. But the study of the comparable corpora of the Brown family yields the observation, in comparing these corpora, that the AmE and BrE often show the same process of change, but that in AmE the change is often ahead that in BrE. For example, the decline of the passive voice and the decline of the modal auxiliaries in the AmE corpora are both about a generation ahead of the BrE corpora. (Figures 5 and 6 – see also Leech et al. (2009).)

For some, this is fairly convincing evidence that AmE, by showing a greater and earlier loss of frequency, is directly influencing BrE. However, others may regard it simply as evidence that AmE and BrE are on the same track, and that AmE happens to be somewhat more advanced, with no assumption of direct transatlantic influence.

The evidence of Americanization is more conclusive if an AmE usage that was virtually non-existent in BrE at time *time-1* becomes established in British usage at time *time-2*, or if at a later time *time-3* the original British usage becomes rare and is supplanted by the American usage. Something like the former can
be observed with the competition between (a) the bare infinitive and (b) the to-infinitive following the verb *help*:  

(a) help solve the problem  
(b) help to solve the problem  

In the 1931 corpus, construction (a) occurred only a handful of times in the British corpus. By 1961 it was getting established, and by 1991 it had increased dramatically, imitating a similar rise in the American 1961 and 1992 corpora. Construction (b), on the other hand, declined slightly from 1961 to 1991 in the British, as well as in the American corpora. The trend, in both cases, was that the BrE corpora were ‘copying’ similar developments in the AmE corpora. This trend
can also be observed in the dramatic reversal, over 1931-1991 in BrE, of the preference for the construction (a) with have as a main verb behaving like an auxiliary over the 'American' construction with do-support (b):  

(a) Have you enough time?  
(b) Do you have enough time?  
(a) He hadn't a chance.  
(b) He didn't have a chance.

(On this topic, see Varela Perez 2007 and Leech et al. 2009: 255-6.)

3.4. Densification

One highly significant trend in both the American and British corpora of the Brown family is an increase in frequency of nouns (see Mair et al. 2002). Since nouns, both as types and as tokens, are the most numerous words in each corpus, and also tend to carry the most information, the most telling outcome of this increased ‘nouniness’ is an increase in density of information per n words. Densification, as this process can be called, is illustrated by a growth in the frequency of content words (especially nouns and adjectives) over function words (such as prepositions and determiners). In part, this can be explained by a growth in frequency of two particular syntactic features of the noun phrase: noun+noun sequences and s-genitives.

The following is an extreme example of how frequency of nouns affects density of information in certain varieties of present-day English writing: the nouns are identified by underlining:

Five syntheses were completed this year and are expected to be published in early 2007: Commercial Motor Vehicle Driver Training Curricula and Delivery Methods and Their Effectiveness; Commercial Motor Vehicle Carrier Safety Management Certification; The Role of Safety Culture in Preventing Commercial Motor Vehicle Crashes; The Impact of Behavior-Based Safety Techniques on Commercial Motor Vehicle Drivers; and Health and Wellness Programs for Commercial Vehicle Drivers. [AmE H06]

The italicized part of the third line shows how juxtaposed nouns can sometimes build up into complex sequences: here there is a sequence of five or six nouns in

---

4 The ratio in favour of (a) was 60:17 in the British 1931 corpus, and this had reversed to 21:126 in the 1991 corpus. The AmE corpora Brown and Frown showed heavy preference for (b), increasing by 1992: from 17:131, to 17:175. See Leech et al. (2009: 255).
a single noun phrase. In the Brown family, the increase of noun+noun sequences is steeper in AmE, but in BrE too, as Figure 7 shows, the increase is remarkable, and is consistent across all four subcorpora.

Figure 7. Increasing frequency of noun+noun sequencies in BrE (1931–1991)

Noun+noun sequences are often more compressed alternatives to constructions using prepositional phrases: for example, *land use* can be replaced by *(the) use of land*, an expression which clearly employs more words and occupies more space. The construction of noun+noun sequences, although by no means a new development in the history of English, appears to have been growing considerably since at least the eighteenth century (Rosenbach 2006), and to have undergone a marked increase since the beginning of the twentieth century, with the growth of the popular press, particularly in the USA. Figure 7 confirms the popularity of this type of construction in journalism, as it shows Press as the subcorpus with the highest frequency of noun+noun sequences, and also shows a continuing marked rise (cf. Biber 2003). Clearly noun+noun sequencing is particularly characteristic of certain genres – not only journalistic, but also technological and bureaucratic. However, it is noticeable from Figure 7 that this feature of noun phrases has increased even in fiction writing, where technical compression of information is less expected.

The increase in the *s*-genitive has also been remarkable over the twentieth century (see Hinrichs and Szmrecsanyi 2007), as Figure 8 shows for BrE:

The increase of *s*-genitives, like that of noun+noun sequences, is steepest in the Press subcorpus, which also registers the highest frequency of this construction. Again, as in the case of noun+noun sequences, the effect of the *s*-genitive is often to compress information into the pre-modifying part of
the noun phrase, rather than to spread it out through the use of prepositions, especially of. Compare the length of the following examples:

the poll’s results ←→ the results of the poll
the nation’s history ←→ the history of the nation

3.5. Prescriptivism

Prescriptivism is not so much a process as a state of mind: an attitude which favours certain usages and rejects others, often without good reason. Such attitudes can be widely and influentially advocated (e.g. by teachers, textbook writers, publishers, language mavens and grammar checkers), so that they can cause change in language behaviour, especially in written English, through the changing attitudes, often accompanied by guilt, they bring about in others. Two examples where prescriptivism appears to be part of the explanation of change are (a) the decline of frequency of the passive (see Figure 5) and (b) the marked change of preference from which to that as a restrictive relativizer, as in (1) and (2):

(1) Tarrant repeated at length the exchanges which had taken place in Scotland Yard. [LOB LO2]
(2) Tarrant repeated at length the exchanges that had taken place in Scotland Yard.

The declining frequency of the passive may be due to a combination of processes: colloquialization, Americanization, and prescriptivism. Since the
passive is far less frequent in conversation than in academic writing (where, as we have seen, conservative standards of writing are most strongly maintained), passive decline can be attributed, at least in part, to colloquialization – the spread of spoken habits of usage to writing. Also, as we have seen, AmE has been leading the way in this decline – so Americanization is also in play. However, it is also well known that powerful prescriptive forces have been advocating the avoidance of the passive during the twentieth century – a trend more recently reinforced by the effect of automatic grammar checking facilities incorporated in word-processing software. The evidence that the decline of the passive has been increasing suggests that such disapproval by ‘language authorities’ is having its effect. It also appears that this anti-passive propaganda is more powerful in the USA than in the UK (see Seoane and Williams 2006) – hence the association of prescriptivism in this case with Americanization.

The second example of prescriptivism also emanates from sources in the USA, where the ‘sacred That Rule’ (as it has been called by Arnold Zwicky) put forward forcefully by ‘language authorities’ decrees that which be avoided as a relativizer in restrictive relative clauses, and that that be used instead (i.e. in the examples above, (2) is prescribed instead of (1)). Here again, colloquialization and Americanization can be linked together as contributing processes, but the most influential factor in causing a switch from wh-relativizers to that (and, to a lesser extent, to zero relativization) appears to be prescriptive pressure. Figures 9 and 10 show the increase of that-relativization and the decline of

\[ \text{Figure 9. Frequency pmw of that-relative clauses in the Brown family of corpora, with comparison between BrE 1931-1991 and AmE 1961-1992.} \]

5 See the discussion of this issue in Zwicky’s ‘Language Log’ (posted 22 May 2005 at http://itre.cis.upenn.edu/~myl/languagelog/archives/002291.html#more.)
wh- relativization in both AmE and BrE. In AmE, the decline of wh- relativizers is largely due to avoidance of which rather than of who, whom or whose. In BrE, although the use of which declines, who, whom and whose also account for part of the decrease.

It appears that BrE, up to 1991, was little affected by the ‘sacred That Rule’, and that the moderate decline of wh- pronouns may have been due more to the negative effect of colloquialization.

4. Conclusion

The foregoing account of change in English grammar highlights some of the most significant findings that have emerged from the study of the Brown family of corpora over the last decade. What I have presented is no more than an outline of ‘work in progress’ which continues as more corpora are added to the family, and as the analysis is extended to new periods. Moreover, the Brown family, being restricted to printed and published resources, needs to be supplemented by other diachronic corpora, especially the Diachronic Corpus of Present-day Spoken English (DCPSE), the Corpus of Contemporary American English (COCA) and the Corpus of Historical American English (COHA). This is not to mention the many corpora that have been developed or are under development representing other regional varieties of English around the world.

Although details of statistical significance have not been included in this account, other published sources (especially Leech et al. 2009) confirm that almost all of the trends illustrated are highly significant according to log likelihood tests.
The research reported here has been undertaken since 1998 with the support of grant-giving bodies (the Arts and Humanities Research Board, the British Academy and the Leverhulme Trust), and with the help of many co-researchers, especially Nicholas Smith and Paul Rayson, to whom I am grateful for continuing discussion and collaboration.
References


