

LANGUAGE AND COMPUTERS:
STUDIES IN PRACTICAL LINGUISTICS

No 55

edited by
Christian Mair
Charles F. Meyer
Nelleke Oostdijk

The Changing Face
of Corpus Linguistics

Edited by
Antoinette Renouf
and
Andrew Kehoe

Rodopi

Amsterdam - New York, NY 2006

Recent grammatical change in written English 1961-1992: some preliminary findings of a comparison of American with British English

Geoffrey Leech and Nicholas Smith

Lancaster University

Abstract

The quartet of corpora analysed in this paper are the Brown Corpus (AmE, 1961), LOB Corpus (BrE, 1961) Frown Corpus (AmE, 1992) and FLOB Corpus (BrE, 1991). The POS-tagged versions of these matching corpora provide the basis for tracking frequency changes in grammatical usage in written English 1961-1991/2 and for comparing similar changes in AmE and BrE. For example, there have been significant increases in the use of semi-modals, the present progressive, that-relativization, nouns (in particular proper nouns), s-genitives, and verb and negative contractions. Counterbalancing some of these changes, there have been significant decreases in the use of core modals, the passive voice, wh-relativization, and of-genitives. In general, the changes in AmE are more extreme than those in BrE. We discuss these changes in terms of general diachronic processes, particularly socially determined processes such as colloquialization and Americanization.

1. Introduction

This paper can be seen as building on the work of Marianne Hundt and Christian Mair in investigating the potential of the matching Brown, Frown, LOB and FLOB corpora for revealing recent changes in English grammar. Hundt (1997), for example, explored differences between certain parts of these matching corpora of written AmE and BrE corpora to show how 'AmE, with the occasional exception, is usually more advanced in ongoing morphological changes'. In this paper we revisit some patterns of change she noted, as well as some additional grammatical topics, taking advantage of the four complete corpora which have now been POS (part-of-speech) tagged.

In recent publications on the Brown family of corpora¹ (Leech, 2003, forthcoming; Smith, 2003a, 2003b) we have presented some grammatical findings from a diachronic comparison, particularly between the LOB and FLOB corpora of British English (sampled from publications in 1961 and 1991 respectively). The changes that can be observed in these *comparable* corpora separated by the period of a generation are changes only of frequency of use, but nevertheless some notable patterns of increase and decrease emerge from their comparison.

Throughout, our focus is on WRITTEN, published English. In most instances, it is likely that any changes observed will have previously been

initiated and propagated in the spoken registers of the language variety. However, the present scarcity of suitable spoken BrE and AmE corpora from the early 1960s prevents us from carrying out an empirical investigation of the relationship between spoken and written language.

The present paper is a progress report on a further extension of this research, funded by the British Academy.² This enabled us³ to use the POS-tagging in comparing the frequency of occurrence of selected grammatical categories in the American corpora (Brown and Frown) as well as in the British corpora of the same dates.⁴ In other words, it was possible to make a four-way comparison, as represented in Figure 1.

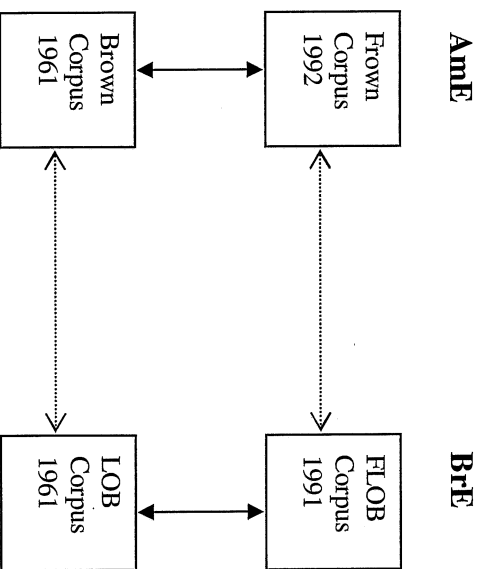


Figure 1: The Brown, Frown, LOB and FLOB corpora

By *comparable corpora* we mean corpora built according to the same principles of design and selection, such as Brown, LOB, Frown and FLOB. In principle, comparisons between such corpora, separated by a period of 30 years in this case, provide a uniquely precise way of tracking historical developments in language use. But the comparisons are by no means free of problems. Leech (2004) attempts to clarify some hazardous assumptions involved in frequency studies of comparable corpora such as those of the Brown Family. They include the following assumptions:

- (a) that the size and composition of the corpora are sufficiently closely matched to validate the basic principle of the comparison: that we are comparing like with like despite different provenances;
- (b) that the statistically significant results of the comparisons can be attributed to linguistic differences rather than other factors such as shifts in genre characteristics;

- (c) that the grammatical categories are defined and used consistently and in a way that other linguists will find useful;
- (d) that the extraction of classified data from the corpus has been acceptably, if not totally, free from error.

In the present state of our American English–British English comparison, which relies on part-of-speech tagging for many of its grammatical comparisons, factor (d) carries an additional hazard. Whereas the British corpora (LOB and FLOB) have been manually post-edited after POS tagging, the same is not true of the American corpora, in which we estimate that an error of 2% in the assignment of POS tags remains. From previous studies of errors in automatic tagging of the LOB and FLOB corpora (see Mair et al, 2002: 262–264), we have been able to arrive at estimates of frequency incorporating a corrective coefficient. But where extraction is reliant on grammatical distinctions as opposed to purely orthographic or lexical ones, there is undeniably a higher degree of approximation in the statistics derived from Brown and Frown than from those derived from LOB and FLOB.

A defence of this and the previously published comparisons, however, is that the differences of frequency are in many cases so highly significant that the results are unlikely to change materially as a result of further work of correction and verification. Nevertheless, the residual element of doubt makes it appropriate to apply the term ‘provisional’ to any of our grammar-dependent findings from comparable corpora. In other words, they are accepted *pro tem* in the expectation that further corpus-based or other empirical research will confirm and refine the findings. While this caveat also applies to the present report, we reiterate the point that many of the frequency findings to be presented are so striking as to allay doubts as to the general descriptive conclusions arrived at. Increasingly, there are also relevant results from other studies⁵ which tend to offer corroboration of these trends.

The findings we present belong to two major units of grammar: to the verb phrase and to the noun phrase. In discussing these findings, we repeatedly find ourselves referring to colloquialization and (for British results) Americanization as likely explanatory factors.

2. Findings concerning categories of the verb phrase

The verb categories we deal with here are those relating to modality, progressive aspect, passive voice and subjunctive mood. These happen to be categories all showing some striking differences as well as similarities between the AmE and the BrE corpora.

2.1 Modal auxiliaries and so-called semi-modals

We have already reported findings on this topic in the publications cited at the beginning of this paper. However, the focus there was not on the AmE – BrE

difference. As Table 1 and Figure 2 illustrate, there has been a decline in the use of the 'core' class of modals *would*, *will*, *can*, *could*, *may*, *should*, *must*, *might*, *shall* plus the marginal modals *ought (to)*, *need* (+ bare infinitive). In the four corpora overall, these 11 modals occur in the order of frequency corresponding to the order just given, and in fact, the order varies very little among the four corpora. But the decline is much steeper in the case of the middle-order members of the list, *may* and *must*, and particularly the bottom-ranking members, *shall*, *ought (to)* and *need*.

Table 1: Frequencies of the core modals in AmE and BrE

	American English			British English		
	Brown (1961)	Frown (1991)	Change %	LOB (1961)	FLOB (1991)	Change %
<i>Would</i>	3,053	2,868	* -5.9%	3,032	2,682	** -11.5%
<i>Will</i>	2,702	2,402	** -11.0%	2,822	2,708	-4.0%
<i>Can</i>	2,193	2,160	-1.4%	2,147	2,213	+3.1%
<i>Could</i>	1,776	1,655	* -6.7%	1,741	1,767	+1.5%
<i>May</i>	1,298	878	** -32.3%	1,338	1,100	** -17.8%
<i>Should</i>	910	787	** -13.4%	1,301	1,148	** -11.8%
<i>Must</i>	1,018	668	** -34.3%	1,147	814	** -29.0%
<i>Might</i>	665	635	-4.4%	779	640	** -17.8%
<i>Shall</i>	267	150	** -43.8%	355	200	** -43.7%
<i>ought (to)</i>	69	49	-28.9%	103	58	** -43.7%
<i>Need</i>	40	35	-12.4%	76	44	** -42.1%
Total	13,991	12,287	** -12.1%	14,841	13,374	** -9.9%

NOTE: The figures in the columns headed by Brown, Frown etc. are frequencies per million word tokens in the corpora; the next column gives the changes in frequency expressed in percentages, i.e. the difference between the two frequencies as a percentage of the first. In addition, a probability value is reported if this is calculated to be statistically significant: * indicates a probability of less than .05, ** a probability of less than .01, and *** a probability of less than .001, of any observed diachronic change. The probability was obtained using the log likelihood test of significance (Dunning, 1993).

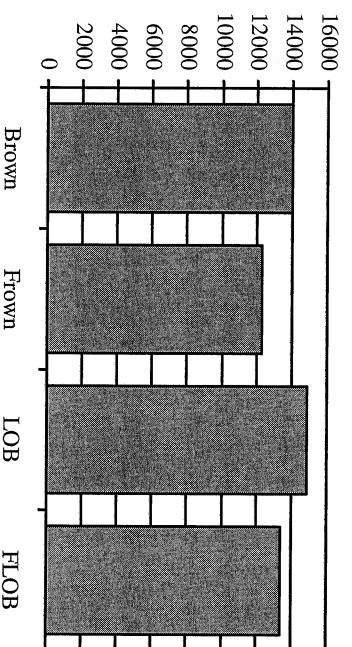


Figure 2: Declining profile of the core modals in AmE and BrE

The frequency decline of individual modals can be observed from Table 1, which for convenience is reprinted – with slight updating – from Table 4 of Leech (2004), except that the AmE figures are given on the left. (It makes sense to place the American figures first, as AmE typically shows a tendency to go further, or move faster, in a particular frequency change than BrE).

The modals show a 'follow-my-leader' pattern, whereby BrE reaches, by 1991, approximately the same frequency pattern as AmE had in 1961. The decline, as shown most graphically in Figure 2, is considerable – in the region of 10% over the 30 year period – though somewhat higher in AmE (12.2%, as compared with 9.5% in BrE).

The decline in frequency of the modals is countered by an appreciable increase in frequency in both the AmE and BrE corpora of the modal verb idioms often termed 'semi-modals'. Some of these have been widely discussed and investigated (e.g. by Biber et al, 1998: 205-210, Krug, 2000) regarding the grammaticalization thesis that a new generation of modal verbs has been emerging in Modern English, and (more cautiously) that these are in some degree displacing the 'core' modals listed in Table 2. The negative side of this thesis is most persuasive in the case of *must* and *HAVE to/NEED to*. (Note that the italic capitals indicate the lemma rather than the base form *have* alone.) These apparently competing forms, together with similar verbal expressions of obligation/necessity, have been investigated in LOB and FLOB by Smith (2003b).

Table 2: Frequencies of some semi-modals in the four written corpora

	American English			British English		
	Brown (1961)	Frown (1991)	Change %	LOB (1961)	FLOB (1991)	Change %
<i>BE going to</i>	216	332	** +53.9%	248	245	-1.5%
<i>BE to</i>	344	217	** -36.8%	451	376	** -16.9%
<i>(had) better</i>	41	34	-17.0%	50	37	-26.2%
<i>(HAVE) got to</i>	45	52	+15.7%	41	27	-34.3%
<i>HAVE to</i>	627	639	+2.0%	757	825	+8.7%
<i>NEED to</i>	69	154	** +123.5%	53	194	** +265.0%
<i>WANT to</i>	323	552	** +71.1%	357	423	* +18.2%
<i>BE supposed to</i>	48	55	+14.7%	22	47	** +113.1%
<i>used to</i>	51	74	* +45.3%	86	97	+12.5%
Total	1,764	2,109	** +19.7%	2,065	2,271	** +9.7%

Nevertheless, the overall picture is less than persuasively in favour of the displacement thesis: the increasing use of the semi-modals, significant though it is, still leaves the 'core' modals overall vastly more frequent in our data. (This is further discussed in Leech, 2003: 235-237 and Leech, 2004.) Part of the explanation, apparently, is that most semi-modals are primarily spoken forms and – in spite of colloquialization – they are still largely avoided in written English.

The category of semi-modals is not well defined.⁶ To avoid any particular bias, we included in this comparison a broad spread of these verbal idioms, some of which have been declining, whereas others have been increasing dramatically. Those apparently declining or at least not increasing overall are *BE to*, *(had) better* and *(HAVE) got to*, while those apparently increasing are *BE going to* (in AmE), *HAVE to*, *NEED to*, *BE supposed to*, *used to*, and *WANT to*.

Some of the more striking results from the diachronic comparison are:

<i>BE going to:</i>	increase of 54% in the AmE corpora (i.e. from Brown to Frown)
<i>NEED to:</i>	increase of 123% in the AmE corpora and of 249% in the BrE corpora
<i>BE supposed to:</i>	increase of 113% in the BrE corpora
<i>used to:</i>	increase of 45% in the AmE corpora and of 13% in the BrE corpora ⁷
<i>WANT to:</i>	increase of 71% in the AmE corpora and of 18% in the BrE corpora.
<i>BE to:</i>	decline of 40% in the AmE corpora and of 17% in the BrE corpora

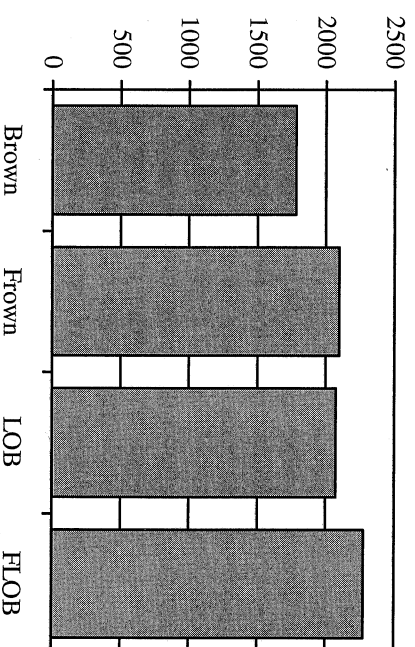


Figure 3: Overall frequencies of semi-modals

Figure 3 shows the overall comparison of the AmE corpora and the BrE corpora in the frequency of occurrence of the listed semi-modals. As a class, they show an increase of 19% in the AmE corpora, and of 10% (very similar to the proportional decrease of core modals) in the BrE. Again, we note a more extreme tendency in AmE. However, surprisingly enough, the semi-modals are overall less frequent in the AmE than in the BrE corpora, so in this respect the picture of AmE beating a path followed by BrE is not maintained.

Giving the overall quantitative results, as we have done, does not reveal any detail as to how or why these changes have been taking place. To dig deeper, we need to observe

(a) How the frequency changes pattern in subcorpora (A-C Press, D-H General Prose, J Learned, K-R Fiction) and in individual text categories (such as Category B – press editorials; Category D – religion; Category K – general fiction).

(b) How the frequency changes pattern in relation to different senses of the modals.

We cannot do more here than highlight two or three of these more detailed observations.

Parallel subdivisions of the corpora yield interesting results in the case of the semi-modals *BE to* and *BE going to* (see also Mair, 1998). The Press section – particularly the category of news reportage – shows a notable decrease in the occurrence of *BE to*, whereas the same section showed a notable increase in the use of *BE going to* (much higher than in BrE generally). As these semi-modals are both associated with future reference, this suggests a switch from the more formal to the less formal (more colloquial) option in the style of news writing – a variety of written language often considered a bellwether for change, and particularly sensitive to changes coming from the spoken language. Limited evidence from spoken corpora (e.g. Leech, 2003: 232) indicates a strong and increasing tendency to use *going to* for future reference in spoken English, so this switch is not surprising. Puzzlingly, though, in other respects the BrE corpora show a slight (non-significant) decline in the use of *be going to* – perhaps a symptom of some resistance to colloquialization in other written genres – especially in the subcorpus of General Prose.

To illustrate changing patterns in the use of modal senses, we examined the frequency of epistemic, root and other senses of the three rather sharply declining auxiliaries *may*, *should* and *must*. We conclude that the root senses of *may* of ‘permission’ (*Please may I finish?*) and of ‘root possibility’ (*as it may be termed*) have been becoming rare in both AmE and BrE, while relatively speaking, the epistemic ‘possibility’ sense has been holding its own and becoming by far the most frequent sense of this modal. With *should*, an opposite trend is observed: the root sense, ‘weakened obligation’, has remained frequent, while the epistemic sense of ‘probability’ and the remaining senses of ‘putative/mandative/quasi-subjunctive’ *should* and *should* as a backshift of *shall* have been becoming increasingly rare, perhaps obsolescent. Unlike *may* and *should*, *must* appears to have declined sharply in both root and epistemic senses. From independent corpus evidence (albeit tentative) – see Leech (2003: 232–233) – these trends are paralleled by similar but somewhat more extreme changes in the spoken language. From these findings, there appears to be some trend towards monosemy accompanying the decline of modals, although no such trend is perceptible with *must*.

2.2 Progressive Aspect

Another verb category of greater frequency in the spoken language is the progressive aspect, which has been broadening its range of application in English since late ME, and is still gaining in frequency (see Mair and Hundt, 1995). In the Brown family overall, the use of the progressive has increased by 11.4%, but the picture is highly variable according to grammatical subcategories. The present progressive active is the most common variant of all, and increases by a

remarkable +31.8% in AmE and +28.9% in BrE. On the other hand, the past progressive declines slightly, by 1.3% (AmE) and 9.0% (BrE). Other areas of pronounced increase are the combinations modal+progressive and (in BrE) progressive+passive. (Oddly, though, the progressive passive undergoes a decline in AmE – part of a general trend discussed in 2.3 – also frequencies of the modal+progressive combinations rise in BrE, whereas in AmE they remain more or less unchanged, at a lower overall level.) The progressive copula – another construction which like the progressive passive is relatively rare and historically rather recent – increases greatly from 3 to 20 in Brown/Frown and from 8 to 17 in LOB/FLOB, although the numbers are too small for statistical significance. Examples include:

- (1) Maybe my friend, who happens to be white, and I *are being too heartless*. Or maybe *we're just being too jealous*. (FROWN, B)
- (2) "You're *being a pain in the neck*," John said. (FROWN, L)

Apart from growth in minor areas such as the progressive copula, detailed examination of the patterns of progressive usage in the LOB/FLOB corpora has failed to reveal any general explanation for the increase of the progressive – for example, extension of its use in certain ‘non-progressive’ classes of stative verbs. It is worth mentioning a probable increase in the occurrence of the so-called interpretative progressive – where the progressive refers to an underlying psychological interpretation of an overt form of behaviour (particularly verbal behaviour). This usage (see Ljung, 1980 and König, 1980, 1995) appears to be one area showing an extension of progressive meaning in recent years. Examples:

- (3) When he speaks of apocalypse, however, he *is not speaking of it* in the literal and popular sense. (FROWN, D)
- (4) *Am I shocking you?* (FROWN, K)

Another atypical usage which shows an increase is the ‘matter-of-course’ use of the progressive with future-referring modals, particularly *will*:

- (5) He *will be standing down* at the next election. (FLOB, B)
- (6) Many of you *will be bringing* your camera along to record the weekend. (FLOB, E)

The effect of using the progressive here (as compared with the non-progressive use of *will*) has been variously explained as (a) disclaiming human intention, (b) expressing a non-immediate consequence of what has already been determined (see Smith, 2003a; Williams, 2002).

2.3 Passive Voice

There is a consistent fall in the frequency of the passive voice, in both BrE (-12.4%)⁸ and in AmE (where it is more extreme at -20.1%).⁹ The declining use of the passive may be considered as another case, like the decline of the modals, of AmE leading the way. However, the situation is different: synchronically speaking, whereas the modals are much more frequent in conversation than in written language, the converse is the case for the passive (Biber et al, 1999: 476). This can, therefore, be seen as a negative example of colloquialization: where the passive, while still strongly entrenched in the more academic varieties of the written language, suffers from a declining popularity consonant with increasing 'oral' influences on writing. Another explanation, however, could be that the sustained attacks on the passive by usage manuals and (most recently) automated grammar checkers have had their effect, especially in AmE.

Table 3: Declining frequency of the passive, by subcorpus

	American English			British English		
	Brown (1961)	Frown (1991)	Change %	LOB (1961)	FLob (1991)	Change %
Press	10,894	7,904	** -27.4%	12,992	11,368	** -12.5%
Gen prose	12,691	10,400	** -18.1%	14,983	13,126	** -12.4%
Learned	19,177	14,180	** -26.1%	20,601	17,183	** -16.6%
Fiction	5,582	5,290	-5.2%	6,113	5,895	-3.6%
OVERALL	11,588	9,254	** -20.1%	13,260	11,614	** -12.4%

As Table 3 shows, the decline in the use of the passive is pervasive in the sense that each subcorpus shows a decline for both AmE and BrE. It is noteworthy, however, that the decline is proportionately high in the Learned subcorpus, where the passive is most frequent, and is low in the Fiction subcorpus, where the passive is least frequent.

2.4 Subjunctive Mood

In the mid-20th century, the subjunctive mood in British English was typically regarded as an obsolete relic of older English, virtually on the brink of extinction. However, by the end of the century a different perspective was being presented: the British use of the mandative subjunctive in *that*-clauses (as exemplified below) was seen to be making a come-back (e.g. Overgaard, 1995). Examples of the mandative subjunctive are:

- (7) The doctors had suggested *Scotty remain* most of every afternoon in bed until he was stronger. (BROWN, K)

- (8) Hence it is important that the process *be* carried out accurately (FLob, H)

The mandative subjunctive, in the four corpora under examination, has indeed undergone a modest revival in BrE: rising from the low figure of 14 occurrences in 1961 (LOB) to 33 occurrences in 1991 (FLob). This corresponds with a decline in the mandative use of *should*, justifiably regarded as the typically British option until recently. In contrast, the figures for AmE show a decline from the relatively high 91 tokens in Brown to 78 tokens in Frown. (However, these figures are too low to be statistically significant and exact exhaustive counts for the later corpora still have to be obtained.)¹⁰

The unusual pattern of reversal of a pre-20th century decline of the subjunctive appears to be a result of American influence on British usage in the (later) 20th century – see further 4 below. Many major grammatical changes seem to be actuated by the growing preference for a more 'oral' style in written language. However, this revival of the mandative subjunctive – a construction associated with formal writing rather than speech – runs counter to the colloquialization trend, and, as Overgaard discusses in some detail, American influence is the only ready explanation.¹¹

3. Findings concerning Aspects of the Noun Phrase

3.1 Relativizers

Relativization constructions show big changes in the four corpora, and again colloquialization seems to be a major factor. In AmE, a dramatic decline of *which* as a relative pronoun (-34.9%) confronts an even more dramatic increase in the use of *that* as a relativiser (+48.3%). BrE shows the same trend, but to a far less extreme extent (-9.5% for *which*, +9.0% for *that*). Other trends in relative clause construction are an increase in the use of zero relativization (AmE +23.1%, BrE +17.1%),¹² together with a decrease in pied-piping (preposition+relative pronoun constructions – of *whom*, in *which* etc.) of -15.9% in AmE and -16.9% in BrE, and a corresponding increase in preposition stranding (+19.5% in AmE and +97.4% in BrE).¹³ Examples of zero relativization, pied-piping, stranding with zero, and stranding with *that*:

- (9) But the seven-iron shot he used to approach the green strayed into a bunker (Brown A)

- (10) (...) this is an area in which Diana is especially interested (FLob G)

- (11) I've found him to be the most casual politician I have ever worked with (Frown B)

- (12) "There's nothing there that you object to?" I asked her. (LOB P)

The overall picture in AmE and BrE is similar: a decrease in *wh*-relativization is balanced against a converse rise in zero and *that* relativization. This is consonant with the colloquialization thesis: the *wh*-forms (proportionately less common in conversation – see Biber *et al* 1999: 610–11) are declining and *that* and zero (the options most used in conversation) are increasing.

Table 4: Decline of *wh*-relativization v. increase of *that* relativization

Relativization type	American English			British English		
	Brown (1961)	Frown (1991)	Change %	LOB (1961)	FLOB (1991)	Change %
<i>wh</i> -rel.	6,034	4,890	** -19.0%	6,971	6,376	** -8.5%
<i>that</i> rel.	1,803	2,674	** +48.3%	1,346	1,467	* +9.0%

A further, connected, sign of colloquialization in relative clauses is the trend towards preposition stranding, alongside the converse decline of pied-piping.

When we look at individual *wh*-pronouns, on the other hand, there is a clear difference between AmE and BrE. In AmE, the decline of *which* alone accounts for all of the loss of *wh*-relativization: in fact, *who*, *whom* and *whose* all increase slightly.¹⁴ In BrE, on the other hand, there is a small but rather consistent decline in all three *wh*-relative pronouns. What stands out here is the increasingly pronounced American disfavour for *which* as a relative pronoun (except in non-restrictive clauses), and the American preference for *that* as an alternative. This preference, amounting to an increasing taboo against *which* as a restrictive relativizer, is now built into grammar checking software, and we can expect it to be making even greater headway at present than in the early 1990s.

3.2 Other Features of the Noun Phrase

3.2.1 Nouns

Turning to other aspects of the noun phrase, at this stage we can only indicate roughly what provisionally observed changes invite further research. According to initial research, nouns as a part of speech have increased their frequency of occurrence by more than 4% in the Brown family of corpora (+4.0% in AmE, +5.3% in BrE). Part of this may be due to an increasing popularity of noun+noun sequences (approximately +10 % in AmE, +17% in BrE), also of proper nouns (+12.8% in AmE, +10.0% in BrE.). This higher frequency of nouns again runs counter to the colloquialization thesis: high noun frequency is associated with high density of information, and is a marked characteristic of informative as contrasted with interactive written styles (Biber, 1988: 89). Biber & Clark (2002) found a similar trend across a wider diachronic span.

3.2.2 Genitives and of

The *of*-construction seems to some extent to be giving way to a more frequent use of the equivalent *s*-genitive construction. According to the Brown family of corpora, the increase of the genitive over the 30-year period is remarkable: +41.9% in AmE, +24.1% in BrE. The decrease in the use of *of* over the same period is less remarkable in percentage terms (–10.6% in AmE, –4.7% in BrE), but bearing in mind the very high frequency of this preposition, the decrease is also highly significant. More relevant, though, is a comparison of genitives with *of*-phrases which are semantically equivalent to genitives, or *of*-genitives as we may conveniently call them – e.g. *the common soldiers' letters* compared with *the letters of the common soldiers*. Taking a 2% sample of each corpus, we arrived at tentative figures of –31.9% for AmE and –23.6% for BrE. The loss of *of*-genitives is very roughly commensurate with the gain of *s*-genitives.

Like the competition between *wh*- and *that* relativization, the competition between *of*-genitives and *s*-genitives fits into the mould of colloquialization.¹⁵ Both *wh*-relativizers and *of*-genitives arose in ME and gained strength apparently through the influence of analogous structures in French and Medieval Latin (Mustanoja, 1960: 78; Fischer, 1992: 301). In the present age, when the spoken medium is asserting itself more powerfully, a resurgence of the *s*-genitive and *th*-/zero relativization, structures which owe nothing to Romance models, appears to be taking place.¹⁶

3.2.3 Personal pronouns

One of the puzzling results of the comparison of these corpora is in the frequency changes of 1st and 2nd person pronouns. The pronoun *I* increases its use by +31.2% in AmE, whereas it decreases its use (–10.1%) in BrE. A similar contrast is seen in the plural pronoun *we*: AmE +12.8%, BrE –6.9%. In fact, both pronouns manifest a cross-over phenomenon whereby the frequency in LOB approximates to that in Frown, and the frequency in FLOB approximates to that in Brown:

<i>I/me/my/mine/myself</i>	
AmE	BrE
(1961) [7,560]	(1961) [7,560]
(1991/2) [7,531]	(1991/2) [6,793]
<i>we/us/our/ours/ourselves</i>	
AmE	BrE
(1961) [3,112]	(1961) [3,112]
(1991/2) [3,162]	(1991/2) [2,889]

You also shows a large increase in AmE (+18.0%), but is virtually unchanged in BrE (+0.2%). On the face of it, the increase in 2nd person and (especially) 1st person pronouns in AmE is another sign of colloquialization: these pronouns are strongly associated with the personal style of communication found in conversation. Thus the changes in AmE make sense in terms of the adoption of a more interpersonal, speech-oriented style of address in the written language. But the absence of such changes in BrE, and even more so the converse trend in the first person pronoun use, are mystifying. Further research is needed.

More easily interpretable are the changes in third person pronoun use: *HE* loses frequency (-22.9% in AmE, -8.8% in BrE), *SHE* gains frequency (+34.9% in AmE, +8.8% in BrE), yet *SHE* is still less frequent than *HE* in the later corpora. It scarcely needs comment that during the 1961-1992 period, when the women's movement had its major impact, female references gained at the expense of male references, and yet male references still predominated over female in the 1990s. It is also unsurprising that in the written language, the use of *HE* as a gender-neutral pronoun declined, and that non-sexist alternatives such as *HE* or *SHE* made an (increased) appearance. But the numbers of occurrences of these composite pronouns are small: the increase in the overall count goes from 9 to 56 in AmE, and from 11 to 37 in BrE. Another solution to the problem of gender bias is the use of 'singular *THEY*' in the sense of 'he or she'. We analysed a sample of 6% of the corpus tokens of *THEY*, i.e. roughly 500 instances from each corpus randomly selected, and found an increase from 7 to 9 occurrences of 'singular *they*' in AmE, and of 0 to 9 occurrences in BrE. From an equivalent sample size of *HE*, we found a reduction from 20 to 7 instances of gender-neutral *HE* in AmE, and from 32 to 4 instances in BrE. These numbers are very small, but if we postulate (speculatively) a scaling-up of these results to each pair of corpora as a whole, they point to a pivotal shift from the use of *HE* as a gender-neutral singular pronoun (a hypothetical decrease from 866 to 183 tokens), to the use of alternatives such as *THEY* and *HE* or *SHE* (the former with a hypothetical increase from 17 to 300 tokens).

4. Conclusion: Colloquialization? Americanization?

Among the frequency changes taking place in the grammar of the noun phrase and the verb phrase between 1961 and 1991/2, it is easy to notice that many show a tendency for spoken language habits to infiltrate the written language: colloquialization. Perhaps the most conspicuous sign of such an infiltration is the increasing inclination to use of verb contractions (*it's*, etc.) and negative contractions (*wouldn't*, etc.) in the four corpora. The overall increase of these contractions is +63% in AmE, and +25% in BrE. Each class of contractions is represented in Table 5:

Table 5: Increasing use of contractions in AmE and BrE

	American English		British English	
	Brown (1961)	Frown (1991)	LOB (1961)	FLOB (1991)
				Change %
verb	2,807	5,032		** +79.3%
contraction			3,126	3,867
neg.	2,087	2,959		** +41.8%
contraction			1,940	2,462
				** +26.9%
Total	4,894	7,991	5,066	6,329
		** +63.3%		** +24.9%

This table also illustrates very clearly another typical trend, whereby AmE shows a more extreme change of frequency than BrE. What is less typical here is that at the starting point (1961) AmE shows a lower frequency than BrE, whereas at the finishing point (1991/2), AmE shows a high frequency than BrE. In other words, AmE seems to have overtaken BrE in the use of contractions during this 30-year period.

Since colloquialization appears to be the 'default' trend, if we find a case where an opposite trend takes place, this invites explanation (and further research). For example, the decidedly uncolloquial mandative subjunctive construction discussed above shows an increase in BrE, but this is a change in an opposite direction to colloquialization. The explanation here appears to be that Americanization – a trend which often goes hand in hand with colloquialization, in this case militates against it. It is as if here the usage imperative 'Adopt a more American style' outweighs the imperative 'Adopt a more colloquial style'.

There is not unreasonable scepticism over the interpretation of terms like 'colloquialization' and 'Americanization', and certainly these are not uniform trends. They are cases where they appear to operate very clearly (as in the data for contractions in Table 5) and there are cases where they don't seem to operate at all – as in the declining frequency of *I* and *we* in the BrE corpora. Perhaps the term 'colloquialization' conceals more than one factor with different effects.

Another argument might be that in a case where AmE 'leads the way' and BrE follows some way behind (as in the case of declining modals), this is not necessarily a case of Americanization: perhaps it is simply that two regional varieties of the same language follow the same course of change, but that the change is more advanced in one variety than the other. Study of other regional varieties (e.g. Australian, Irish) might throw further light on this. No causative influence of one variety on the other need be implied.

On the other hand, the influence of American usage (like other pervasive American cultural influences we are familiar with) is clear enough, if we examine lexical changes of frequency in the four corpora. For example, *movie(s)* is a noun which in 1961 was almost confined to AmE (67 occurrences in Brown, only 7 in LOB); now it has been catching on in BrE, and this shows up in its increasing frequency in the FLOB corpus (120 occurrences in Frown, 35 in LOB). Another characteristically AmE noun is *guy(s)*, which shows a similar trend: (68

occurrences in Brown, only 6 in LOB; 131 occurrences in Frown, 40 in FLOB). In both these cases the noun has nearly doubled its frequency of occurrence in AmE in 1961-1992, but the increase in BrE is five-to-sevenfold.

But again the trend is not uniform: we find cases where AmE and BrE seem to follow diametrically different paths. The progressive passive (see 2.3) becomes more infrequent in AmE and more frequent in BrE. This is a tantalizing case where competing pressures in the two varieties seem to produce opposite results. The progressive passive is a combination of the progressive (which has been gaining frequency) and the passive (which has been losing frequency). It is also suffers from the double-*BE* phenomenon, and as a relatively late historical arrival in English (dating from the late 18th century) is less thoroughly established than other combinations such as the modal progressive and the perfect progressive. These observations draw attention to the uneasy status of the progressive passive, which might lead it to be inhibited in one variety but not in another. It seems that on the present evidence, AmE, with its more pronounced antipathy to the passive, has been affected by this inhibition whereas BrE has not, but has instead followed the trend of greater use of the progressive in this construction, as elsewhere.

There is much need for extensive further research, using both the corpora of the Brown family which have been the focus of this study, and other sources of data, before the tentative claims of this paper can be confirmed. Regarding the latter, our current strategy is to extend the Brown family of corpora to earlier points of time. We are currently preparing a corpus of 1930s British English as a mirror to the LOB and FLOB. Results from such corpora should help us to see changing patterns of grammatical use in a wider historical perspective.

Notes

1 We are using this as a convenient term for Brown, LOB, Frown and FLOB. There is no need to go further, and to consider Frown and FLOB as the children, etc. Pursuing the metaphor, the Brown family might be expanded to include collateral kin such as the Kolhapur corpus of Indian English (Shastri, 1988) and Australian Corpus of English (Collins and Peters, 1988), which are matching corpora in design and sampling, but not of comparable dates to Brown and LOB or Frown and FLOB. We do not consider them here.

2 Thanks are due to the British Academy for providing a research grant for this investigation, and also to the Arts and Humanities Research Board for funding that permitted the POS tagging and comparison of the LOB and FLOB Corpora.

3 We are grateful to Christian Mair (University of Freiburg) and Marianne Hundt (University of Heidelberg) for collaboration in the POS tagging of

1 FLOB, and on other aspects of the work on this corpus. Note that Brown was retagged, and Frown was tagged, using the same tagset and tagger as were used for LOB and FLOB, so that grammatical comparisons could be made on the basis of matching grammatical categories.

4 Although the date of Frown text samples is 1992, we assume this is near enough to 1991 to make little difference to the validity of the comparison.

5 See, for example, Övergaard (1995) on the subjunctive.

6 Cf. Quirk *et al* (1985: 136-148), where a cline or gradient between auxiliaries and main verbs is described.

7 Clearly *used to* is aspectual rather than modal in meaning, but it is included here because it is structurally closely parallel to the modals.

8 For convenience, we will henceforth use the minus sign with percentages to indicate a percentage decrease, and the plus sign with percentages to indicate a percentage increase.

9 This decline cannot be attributed to the rise of the *get*-passive, which (with 54 and 72 occurrences respectively in Brown and Frown) is still too rare in the written corpora to have any impact on the use of the standard *be*-passive.

10 Counts are provided by Serpollet (2003) but based on template searches in XKwic which are probably slight underestimates. Compare Övergaard's frequency data (2000: 14-35).

11 Övergaard's summary (1995: 54) is worth quoting: "The distribution of the subjunctive variants in mandative sentences in BrE has changed dramatically during the second half of the twentieth century. What appeared to be a unidirectional drift from the non-inflected morphological variant to the periphrastic variant has not only stopped; we are witnessing a reversal of the drift resulting in increased use of the older non-inflected subjunctive, no doubt due to American influence."

12 These percentages are based on limited sampling: categories A-C for AmE and categories A-E for BrE.

13 However, these percentages are derived from very limited samples, and have been taken as merely indicative of what an exhaustive analysis of the corpora would show.

- 14 The increase in *whom* is more than slight: going from 140 to 165 (+17.5%). This is another surprising change, swimming against the tide of colloquialization, this time in AmE. However, the increase is not statistically significant.
- 15 *Of* is more characteristic of formal written genres than is the *s*-genitive. In the Brown family of corpora, it is most frequent in genres D, H and J, and least frequent in N and P.
- 16 Noun-noun sequences (see the discussion of noun frequency) are, like *s*-genitives, an example of resurgence of native syntactic patterns. Leonard (1968), cited in Leonard (1984: 4), reports that there has been a 'great increase in the occurrences of noun sequences in prose fiction from 1750 to the present day.'

References

- Biber, D. and V. Clark (2002), 'Historical shifts in modification patterns with complex noun phrase structures: how long can you go without a verb?', in: T. Fanege, J. Pérez-Guerra and M. José López-Conso (eds.), *English historical syntax and morphology: Selected papers from IICEHL*, 43–66. Amsterdam: John Benjamins.
- Biber, D., S. Johansson, G. Leech, S. Conrad. and E. Finegan (1999), *Longman grammar of spoken and written English*. London: Longman.
- Blake, N. (ed) (1992), *The Cambridge History of the English Language*. Volume II: 1066–1476.
- Collins, P. and P. Peters (1988), 'The Australian corpus project', in: M. Kytiö, O. Ihalainen and M. Rissanen (eds.) *Corpus linguistics, hard and soft*, 103–120. Amsterdam, Rodopi.
- Dunning, T. (1993), 'Accurate methods for the statistics of surprise and coincidence', in: *Computational Linguistics* 19.1, 61–74.
- Fischer, O. (1992) 'Chapter 4: Syntax', in: Blake (1992), 207–408.
- Hundt, M. (1997), 'Has British English been catching up with American English: in the past 30 years?', in: M. Ljung (ed.) *Corpus-based studies in English: Papers from the seventeenth International Conference on English Language Research on Computerized Corpora (ICAME 17) Stockholm, May 15–19, 1996*. Amsterdam: Rodopi. 135–152.
- König, E. (1980), 'On the context-dependence of the progressive form in English', in: Rohrer, C. (ed) *Time, Tense and Quantifiers: Proceedings of the Stuttgart Conference on the Logic of Tense and Quantification*. Tübingen: Max Niemeyer.
- König, E. (1995), 'He is being obscure: non-verbal predication and the progressive', in: Bertinetto et al. (eds.) *Temporal Reference, Aspect and Actuality*, 2 vols. Torino: Rosenberg and Sellier, pp. 155–168
- Krug, M. (2000), *Emerging English modals: a corpus-based study of grammaticalization*. Berlin/New York: Mouton de Gruyter.
- Ljung, M. (1980), *Reflections on the English Progressive*. Gothenburg Studies in English 46. Gothenburg: Acta Universitatis Gothoburgensis.
- Leech, G. (2003), 'Modality on the move: The English modal auxiliaries 1961–1992', in: Facchinetti, R., M. Krug and F.R. Palmer *Modality in Contemporary English*, Berlin: Mouton de Gruyter, 223–240.
- Leech, G. (2004), 'Recent grammatical change in English: data, description, theory', in B. Altenberg and K. Aijmer (eds), *Advances in corpus linguistics: Proceedings of the 23rd ICAME Conference, Gothenburg, 2002*. Amsterdam: Rodopi.
- Leonard, R. (1968), *The types and currency of noun + noun sequences in prose usage 1750–1950*. Unpublished M.Phil. thesis, University of London.
- Leonard, R. (1984), *The Interpretation of English Noun Sequences on the Computer*. Amsterdam: North Holland.
- Mair, C. (1997), 'The spread of the going-to-future in written English: a corpus-based investigation into language change in progress', in: R. Hickey and St. Puppel (eds.) *Language history and language modelling: a Festschrift for Jacek Fisiak on his 60th birthday*. Berlin: Mouton de Gruyter, 1537–1543.
- Mair, C. and Hundt, M. (1995), 'Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress', in: *Zeitschrift für Anglistik und Amerikanistik*, 43, 111–122.
- Mair, C., M. Hundt, G. Leech and N. Smith (2002), 'Short term diachronic shifts in part-of-speech frequencies: A comparison of the tagged LOB and F-LOB corpora', in: *International Journal of Corpus Linguistics*, 7: 2, 245–264.
- Mustanoja, T. (1960), *Middle English Syntax Part I: Parts of Speech*. Helsinki: Société Néophilologique.
- Övergård, G. (1995), *The mandative subjunctive in American and British English in the 20th century*. Stockholm: Almqvist & Wiksell International.
- Peters, P., P. Collins and A. Smith (eds.) (2002), *New frontiers of corpus research: Papers from the Twenty First International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodopi.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1985), *A comprehensive grammar of the English language*. London: Longman.
- Rayson, P., A. Wilson, and G. Leech (2002), 'Grammatical word class variation with the British National Corpus Sampler', in: Peters et al. (2002), 295–306.
- Seoane-Posse, E. (2002), *On the evolution of scientific American and British English, with special reference to recent and ongoing changes in the use of the passive voice*. Paper presented at the 12th International Conference on English Historical Linguistics, University of Glasgow, August 21–26, 2002.

- Serpellet, N. (2003), *Should and the subjunctive: a corpus-based approach to mandative constructions in English and French*. Unpublished PhD thesis. Lancaster University.
- Shastri, S. V. (1988), 'The Kolhapur Corpus of Indian English and work done on its basis so far', in: *ICAME Journal* 12: 15-26.
- Smith, N. (2002), 'Ever moving on? The progressive in recent British English', in: Peters et al (2002), 317-330..
- Smith, N. (2003a), 'A quirky progressive? A corpus-based exploration of the will + be + -ing construction in recent and present day British English', in: D. Archer, P. Rayson, A. Wilson and T. McEnery (eds), *Proceedings of the Corpus Linguistics 2003 Conference*. Lancaster University: UCREL Technical Papers Vol. 16, 714-723.
- Smith, N. (2003b), 'Changes in modals and semi-modals of strong obligation and epistemic necessity in recent British English', in: R. Facchinetti, M. Krug and F. R. Palmer, *Modality in Contemporary English*, Berlin: Mouton de Gruyter, 241-266.

Social variation in the use of apology formulae in the British National Corpus

Matz Deutschmann

Mid-Sweden University

Abstract

This paper explores sociolinguistic variation in the act of apologising in the spoken part of the British National Corpus. The starting point for the investigation is the 'apology formula', as exemplified by the lexemes 'afraid', 'apologise', 'apology', 'excuse', 'forgive', 'pardon', 'regret' and 'sorry'. The sub-corpus used for the study comprises a spoken text collection of about five million words and represents dialogue produced by more than 1,700 speakers in a number of different conversational settings. More than 3,000 examples of apologising form the basis for the analysis. In the BNC, young and middle-class speakers favoured the use of the apology form. Only minor gender differences in apologising were apparent. The study implies that formulaic politeness is an important linguistic marker of social class and also shows that corpus linguistic methodology can successfully be used in socio-pragmatic research.¹

1. Introduction

This paper will conduct a socio-pragmatic investigation of the apology using a corpus-based methodology. Most previous investigations of this speech act have been based on inauthentic data elicited during role-play situations or discourse completion tests.² Arguably, such data, obtained by asking someone how they think they would react in a given situation, is not likely to coincide with responses produced in 'real life' situations. Discussing the advantages of observational methodology in pragmatic research, Wolfson, Marmor and Jones (1989:194) claim that:

...our own intuitions cannot provide us with a complete picture of the social circumstances that result in a given speech act. It is only through an iterative process which makes use both of systematic observation and increasingly sensitive elicitation procedures that we can begin to capture the social knowledge that is the unconscious possession of every member of a speech community.

Ethnographic approaches have been used in a limited number of apology studies, but for practical reasons most of these have relied on limited data from single genres or very few respondents, and/or have based conclusions on retrospective self- or second-hand reports.³