

## Recent changes in the use of the progressive construction in English\*

Bas Aarts, Jo Close and Sean Wallis

University College London

### 1 Introduction

A classical distinction has entrenched itself in linguistics, namely the diachronic and synchronic ways of studying a language. The first considers language in its stages of development, whereas the latter looks at languages viewed from the present moment. This old Saussurean dichotomy has recently been called into question, and it has been argued that the distinction is artificial (see, for example, Labov 1972). Instead, it is argued that languages change all the time, even within the synchronic phases. As a result of these new attitudes to language development there has emerged a new research impetus in linguistics which concerns itself with what has been called *recent (or current) change* (see Mair 1995, 1997; Mair and Hundt 1995, 1997, Denison 1998, Leech 2000, Smith and Leech 2001, Smith 2005, Mair and Leech 2006, Leech *et al.* 2009). Christian Mair at Freiburg was the first to construct parallel corpora of written British and American English spanning four decades in the twentieth century (the *LOB/FLOB* and *Brown/Frown* corpora). These are excellent resources enabling linguists to research changes in written English over 30 years. Manual searches are still unavoidable, however, as these corpora have not been parsed.

At the Survey of English Usage we have taken Christian Mair's initiative further by constructing a corpus of British English comprising selections of spontaneous spoken English from the *London-Lund Corpus* (dating from the late 1950s to early 1970s) and from the British component of the *International Corpus of English, ICE-GB* (dating from the 1990s). This corpus, which we have called a *Diachronic Corpus of Present-Day Spoken English (DCPSE)*, will allow researchers to investigate recent changes in the grammar and usage of Present-Day English over a period of 25-35 years. DCPSE differs from FLOB and Frown in a number of important ways. Firstly, the corpus is unique in containing exclusively spoken English. We opted for a corpus of spoken English because it is generally recognised that spoken language is primary, and the first locus of changes in lexis and grammar. Secondly, the corpus is parsed, which will permit research into synchronic and diachronic grammatical variation. Thirdly, the corpus is fully searchable using ICECUP, the corpus exploration software that we developed for ICE-GB. We envisage that DCPSE will be a major new resource complementing the Freiburg corpora.

In this paper we will look at the changing use of a particular grammatical construction in English, namely the progressive, which has recently been receiving a lot of attention. Our data are derived from DCPSE. We will show how it can be used to perform constructional searches in spoken English.

## 2 Changes in the Use of the English Progressive: Previous Studies

It is commonly accepted that the progressive increased in frequency during the nineteenth century (see e.g. Denison 1998, Hundt 2004, Smitherberg 2005, Núñez-Pertejo 2007, and Aarts, López-Couso and Méndez-Naya, forthcoming). Recent research has shown that the nineteenth century trend of an increase in the frequency of use of the progressive has persisted into the twentieth century. Hundt (2004) uses ARCHER (A Representative Corpus of Historical English Registers) to track the use of the progressive from 1650 to 1990.<sup>1</sup> Her results indicate a rise in the frequency of the progressive in the twentieth century (lower line in Figure 1).

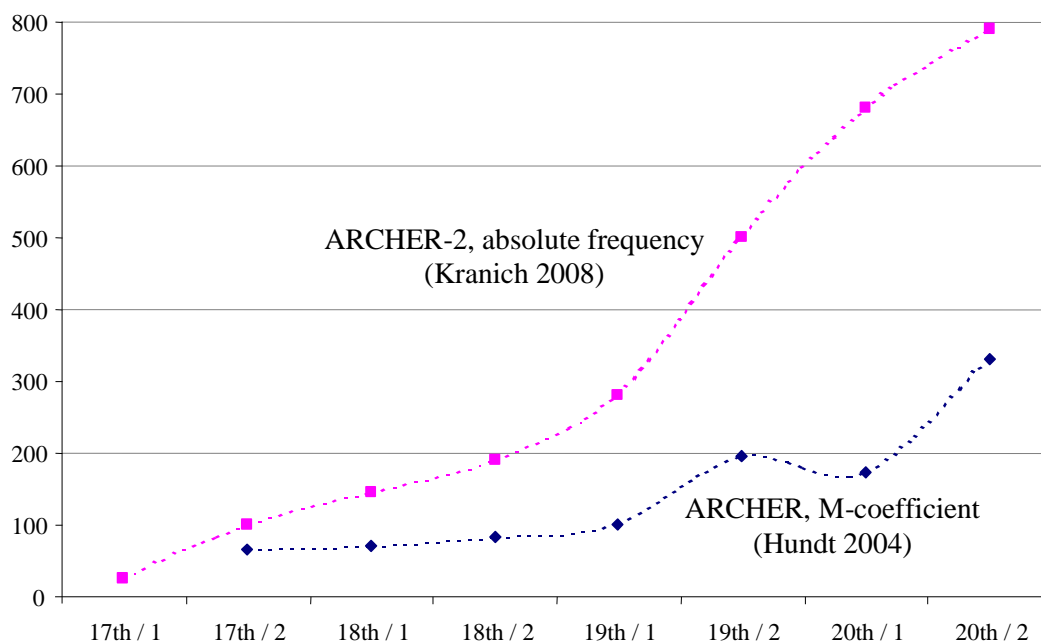


Figure 1: Evidence for the rise of the progressive in Modern British English writing.

Kranich (2008) investigates the progressive using ARCHER-2. Like Hundt, her results indicate a continued increase in the frequency of the progressive in the 20<sup>th</sup> century, as shown in the upper line in Figure 1 (Kranich: 2008: 178).<sup>2</sup> However, what is not clear is whether the rise that is observed is due to a shift toward the progressive within a set of alternative verbal constructions.

Mair and Leech (2006: 323) investigate the increased use of the progressive using the Brown quartet of corpora. Table 1 shows that in British English the use of the progressive seems to be advancing more quickly than in American English. British English has a higher frequency of

progressives than American English in 1961 and the use of the progressive increases by a larger percentage between 1961 and 1991/92.

(PRESS)	1961	1991/92	% rise from 1961
British English (LOB/ FLOB)	606	716	+18.2%
American English (Brown/Frown)	593	663	+11.8%

Table 1: Progressive forms in the press sections (A-C) of four reference corpora LOB, FLOB, Brown and Frown. From Mair and Leech (2006: 323).<sup>3</sup>

Overall this research points to an increased use of the progressive in present-day English, a trend that has continued from the nineteenth century.<sup>4</sup>

Our aim in this paper is to contribute to work on the progressive by investigating the construction systematically in spoken English using DCPSE.

### 3 Exploring DCPSE to Research the Use of the English Progressive

The *Diachronic Corpus of Present-Day Spoken English* (DCPSE) allows us to monitor the use of the English progressive in spoken English over a number of decades. Before showing how this can be done we will introduce a few general features of the corpus and discuss more precisely how the relative frequency of the progressive should be defined.

#### 3.1 The Diachronic Corpus of Present-Day Spoken English

DCPSE was released by the Survey of English Usage (SEU) in 2006. It contains 400,000 words of 1960s spoken material from the *London-Lund Corpus* (LLC),<sup>5</sup> and 400,000 words of 1990s spoken material from the *British Component of the International Corpus of English* (ICE-GB)<sup>6</sup> in matching text categories. DCPSE includes a wide range of spoken English, such as face-to-face conversations, telephone conversations, various types of discussions and debates, legal cross-examinations, business transactions, speeches and interviews. Much of DCPSE is spontaneous, which is important because changes in English propagate themselves in the first instance through spontaneous discourse. It is possible for researchers to listen to the spoken material. As in the previously released *British Component of the International Corpus of English* (ICE-GB), every sentence in DCPSE is syntactically annotated with a phrase structure tree diagram, like the one shown in Figure 2 below:

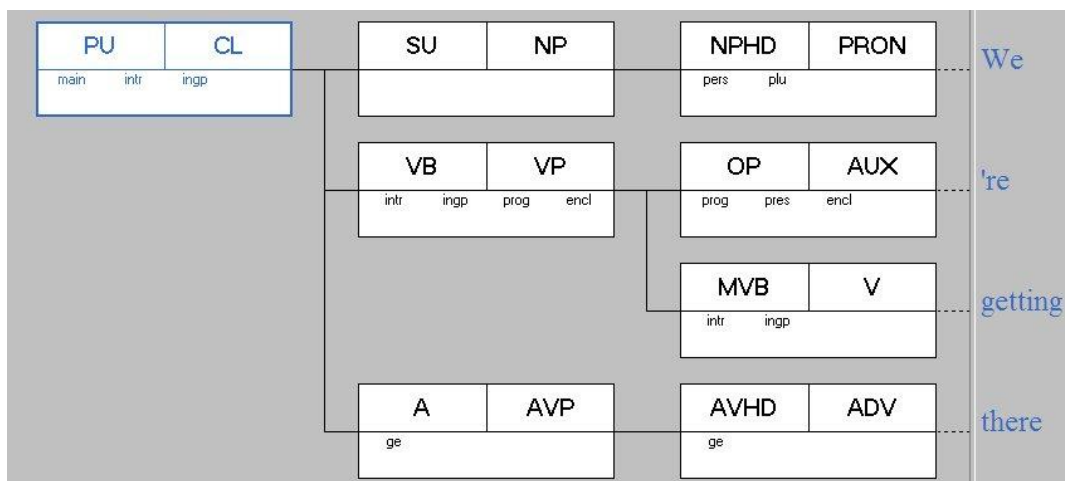


Figure 2: The tree diagram for *We're getting there* from DCPSE (here visualised left-to-right).<sup>7</sup>

In this tree diagram each lexical item, phrase and clause is associated with a node which contains function information (top left), form information (top right), as well as features (bottom portion). Using this architecture DCPSE can be searched with the corpus exploration software ICECUP (the *International Corpus of English Corpus Utility Program*), developed at the SEU. This software enables linguists to search for lexical items and grammatical patterns. ICECUP supports *Fuzzy Tree Fragment* (FTF) queries which allow users to construct approximate (hence 'fuzzy') models of tree structures to retrieve matching cases in the corpus. Figure 3 shows an example of an FTF which retrieves all instances of a VP immediately followed by a direct object (OD).<sup>8</sup>

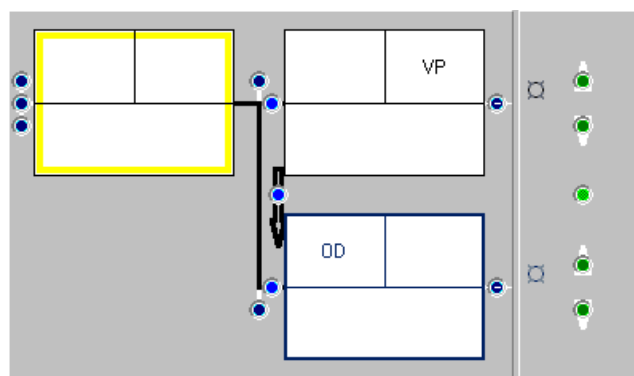


Figure 3: An FTF for a VP followed by a direct object, created with ICECUP.

This brief overview describes only a small amount of the rich functionality of ICECUP: it offers an enormous range of search options, which space limitations do not allow us to describe in detail. DCPSE is an unparalleled resource for linguists interested in short-term changes in spoken English, and is already being used by research groups and doctoral students.<sup>9</sup>

### 3.2 Investigating How the Use of the Progressive Varies

There are a number of ways in which the frequency of the progressive can be calculated in a diachronic study such as this one. A variationist study of the progressive would measure its frequency of use against the frequency of any grammatical variants. The obvious variant to the progressive is, of course, the simple form, although it is clear that the progressive cannot replace any verb phrase in the simple form.

In corpus studies like those by Nehls (1988) and Smith (2002), the progressive is calculated per 100,000 or per million words (using the ‘M-coefficient’, Mossé 1938). Calculating the frequency of the progressive in this way is simple, but it has a major flaw: it does not take into account the possibility that the number of verb phrases per 100,000 words may not be stable diachronically. Nor does it guarantee that the *opportunity* for a progressive to be used is uniform, i.e. as compared to a baseline of ‘progressives plus alternate variants’, as noted in Wallis (2003). We need to treat the graph in Figure 1 with some caution as it shows the variation of progressives in each sample. Strictly speaking we do not know whether this increase is due to a rise in progressive use *per se* or rather that the *opportunity* to use the progressive simply arose more frequently in the data from later centuries.

Smutterberg (2005) discusses this issue and a range of others in his study of the progressive in nineteenth century English and he compares the frequency of progressive use using the M-coefficient and his own ‘S-coefficient’. This S-coefficient is a formula which calculates the number of finite progressives as a proportion of finite verb phrases (excluding the *going to* future and what he refers to as ‘knock-out’ factors, i.e. contexts where the progressive cannot appear).

Smutterberg demonstrates how the chosen methodology affects the results: the S-coefficient shows the progressive has increased by 81% between periods 1 and 3 (1800-1900), while using the M-coefficient puts the increase at 71% (Smutterberg 2005:62). He also found that if the M-coefficient is used, the progressive is most common in Drama, followed by Letters, Trials, Fiction, History, Debates and Science; but if the S-coefficient is used, the progressive is more common in Letters than in Drama, and more frequent in History than in Fiction (Smutterberg 2005: 77-78). In conclusion, simple normalised frequencies are potentially misleading, particularly in the case of differing genres (where the *possibility* of progressives being used may vary) and small samples.

Smutterberg (2005: 46) lists a number of ‘knock-out factors’ in calculating the progressive; these include demonstrations (*I take this hat*), performatives (*I name this ship Elizabeth*), simple imperatives, non-finite verb phrases and stative situations. As discussed by Smutterberg, some of these factors are easier to exclude than others. Imperatives, for example, can be easily removed from any corpus which is tagged, whereas removing stative verb phrases requires manual checking of each example, a time-consuming process.

In calculating the use of the progressive in DCPSE, we follow Smitterberg (2005) in measuring its use against the number of verb phrases, taking knock-out factors into account. As Smitterberg’s study was based on nineteenth century English, some modifications are made. Firstly, we have not excluded stative verbs from the study; Mair and Leech (2006:324) point out that in twentieth century English the progressive may occur with stative verbs, although occurrences are too infrequent to account for the statistically significant overall increase of the progressive. Secondly, in order to exclude demonstrations and performatives as Smitterberg does, each example would need to be manually checked. As they are rare and unlikely to affect the results, they have not been removed.<sup>10</sup>

### 3.3 The Progressive in DCPSE

We used FTFs to look for progressives in DCPSE. The FTF below instructs the search engine to search for a progressive VP (note the feature ‘progressive’ in the bottom section).



Figure 4: An FTF used to search for progressive VPs.

Studies using corpora that are not grammatically parsed must, at this stage, remove instances of the *going to* future (see, for example Smitterberg 2005). However, the *going to* future is not marked as progressive in DCPSE and so is automatically excluded. Our results are summarised in Table 2.

(spoken)	<b>progressive</b>	<b>not progressive</b>	<b>Total</b>	$\chi^2(\text{prog})$
LLC (1960s-70s)	2,399 (3.85%)	59,868	62,267	94.65
ICE-GB (1991/92)	3,152 (5.64%)	52,743	55,895	105.44
<b>Total</b>	<b>5,551</b>	<b>112,611</b>	<b>118,162</b>	<b>200.08s</b>

Table 2: Changes in the proportion of progressive VPs in the LLC and ICE-GB components of DCPSE

The table shows that in the LLC portion, out of a total of 62,267 verb phrases that could have been ‘progressivised’, 2,399 were progressive (3.85%), while in the ICE-GB part, out of 52,743 verb phrases, 3,152 were progressive (5.64%). Formally, we can say that we have refuted the null hypothesis that speakers did *not* change their behaviour regarding the use of the progressive between the two periods.<sup>11</sup>

Comparing the spoken content in the ICE-GB and LLC corpora with the written LOB and FLOB corpora, Smith (2005) found that progressives were almost twice as frequent in spoken rather than written English over the same period (Table 3).

(spoken)	progressive	per million words
LLC (1960s-70s)	2,396	5,990
ICE-GB (1990s)	3,153	7,882
(written)		
LOB (1961)	2,932	2,916
FLOB (1990s)	3,202	3,176

Table 3: Spoken and written language compared (from Smith 2005).

Using DCPSE we can obtain the incidence of the progressive over time. The result is shown in Table 4. Since the total number of cases per year is relatively low, to be on the safe side we also calculate the *Binomial confidence interval* ('error' column).<sup>12</sup> The best way of understanding this error estimate is by looking at a row in the table. In the first sample from 1958, of 1,755 cases of possible progressive VPs, 42 (2.39%) were progressive with an error of  $\pm 0.72\%$  (i.e. from 1.68% to 3.11%). As can be seen by looking at the next row for 1959, the margin of error is greater. As a general rule, the fewer the total number of cases for any year, the larger the margin of error and the

<b>Year</b>	<b>progressive</b>	<b>error</b>	<b>not progressive</b>	<b>Total</b>
1958	42 (2.39%)	$\pm 0.72\%$	1,713	1,755
1959	41 (5.27%)	$\pm 1.57\%$	737	778
1960	112 (4.93%)	$\pm 0.89\%$	2,159	2,271
1961	187 (4.01%)	$\pm 0.56\%$	4,471	4,658
1963	20 (3.90%)	$\pm 1.68\%$	493	513
1964	82 (3.63%)	$\pm 0.77\%$	2,176	2,258
1965	108 (3.93%)	$\pm 0.73\%$	2,643	2,751
1966	123 (3.76%)	$\pm 0.65\%$	3,145	3,268
1967	89 (3.83%)	$\pm 0.78\%$	2,235	2,324
1969	99 (3.19%)	$\pm 0.62\%$	3,001	3,100
1970	80 (5.03%)	$\pm 1.08\%$	1,509	1,589
1971	220 (3.83%)	$\pm 0.50\%$	5,523	5,743
1972	87 (2.85%)	$\pm 0.59\%$	2,965	3,052
1973	37 (5.21%)	$\pm 1.63\%$	673	710
1974	279 (3.90%)	$\pm 0.45\%$	6,874	7,153
1975	561 (4.01%)	$\pm 0.32\%$	13,440	14,001
1976	196 (3.61%)	$\pm 0.50\%$	5,230	5,426
1977	36 (3.93%)	$\pm 1.26\%$	881	917
1990	261 (4.79%)	$\pm 0.57\%$	5,193	5,454
1991	2,193 (5.59%)	$\pm 0.23\%$	37,009	39,202
1992	698 (6.21%)	$\pm 0.45\%$	10,541	11,239

Table 4: Use of the progressive in DCPSE over time.

greater the likelihood that our progressive percentage is not representative of the population.

To make sense of the trend it is helpful to plot this distribution on a timeline. The graph in Figure 5 shows annual data points with the 95% confidence interval expressed as a T-shaped error bar. The graph illustrates the fact that DCPSE does not include data for the period between 1978 and 1989, but this does not prevent us estimating a trend (dotted line). We have also added the centre points (indicated by the 'X' symbols) for the LLC and ICE-GB subcorpora from Table 2.<sup>13</sup>

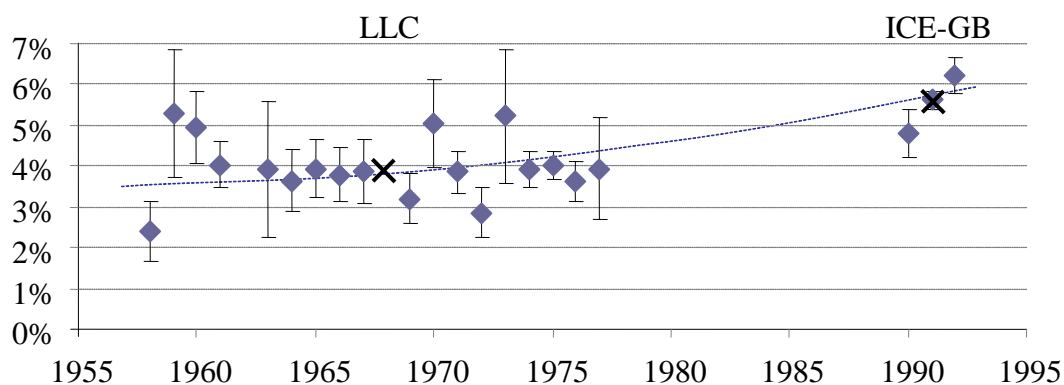


Figure 5: Charting the rise in spoken progressive use in English using DCPSE

#### 4 Why Has the Progressive Increased in Use?

Mair (2006: 88-89) comments that there are three types of changes affecting the progressive:

- (i) many uses which were fully established around 1900 have increased in frequency since then;
- (ii) new forms have been created; and
- (iii) there is a tendency to use the progressive with stative verbs such as *understand*.

The following factors have been suggested by Smith (2005) as probable causes of the increase in the use of the progressive in recent times.

- (i) Contact – the progressive is more common in American English than in British English (Biber *et al.* 1999: 462) and the growing contact between the two countries may have contributed to the increased usage in British English.
- (ii) Increased functional load – Smith (2005: 2) suggests that ‘the progressive has evolved historically such as to convey a rather complex meaning, or set of meanings’ and ‘probably as a result of the varied and developing nature of its meanings, the progressive has enjoyed a meteoric increase in frequency of use’.

Regarding an increase in functional load, Nesselhauf (2007) studied the ‘progressive futurate’ (e.g. *She is graduating next week*) and observed that it tripled in usage between 1750 and 1990. Other linguists, Wright (1994, 1995), Smith (2005) and Smitterberg (2005) among them, have suggested that the so-called ‘interpretive’, ‘explanatory’ or ‘modal’ progressive, as in (1) and (2) below, has



also led to an increase in frequency of the progressive construction in British English, particularly the present progressive.

- (1) If John says that, he's lying.
- (2) When I said the 'boss', I was referring to you. (Huddleston & Pullum 2002: 165).

This use of the progressive 'interprets the speaker's attitude and perspective of the situation; and, in so doing, conveys her epistemic stance at a particular moment in the context of utterance' (Wright 1995: 157). As Smith (2005: 166) puts it: 'Interpretatives are often considered to signal a higher degree of pragmatic meaning and/or subjectivity on the part of the speaker than regular uses of the progressive.' For Quirk *et al.* (1985: 198 fn. b) "the event described has an interrelationship or identity with another simultaneous event", and Leech (2004: 22) observes that "it is as if we are seeing the speech act 'from the inside', not in a temporal sense, but in the sense of discovering its underlying interpretation.". Huddleston and Pullum *et al.* (2002: 165) note that 'regular' progressives in English trigger what they call a 'mid-interval implicature'. Thus if I say *I was reading a book* in answer to the question *What were you doing last night?* there is an implicature that my reading of the book was not coextensive with the beginning and end of the evening. In other words, with such progressives there is an implicature that the time referred to by the progressive is part of a larger situation. This inference is not an entailment, because it is cancellable. Huddleston and Pullum *et al.* (*ibid.*) observe that when the interpretive progressive is used the mid-interval implicature is always cancelled.

Smitherberg (2005: 222) records 364 examples of interpretive progressives over the 19<sup>th</sup> century as a whole in his corpus of 1 million words, with absolute frequencies steadily increasing over his three sub-periods. The interpretive use has also increased during the twentieth century according to Smith (2005), as Table 8 shows.

Genre	LOB (1961)		FLOB (1991)	
	frequency	per million	frequency	per million
Press	9	51	19	107
General prose	23	55	40	97
Learned	0	0	7	44
Fiction	20	78	31	121
Total	52	(52)	97	(96)

Table 8: Genre distribution of the interpretive use of the progressive (present tense), in LOB and FLOB, based on the clearest cases (Smith 2005: 196). Numbers in brackets are means.

A word of caution is in order at this point. In our view it would be prudent not to draw too firm conclusions from the data in the various sources we discussed regarding the increased use of the interpretive progressive, given the differences between scholars' definitions of the concept and the difficulty of detecting an interpretive meaning in individual examples.

Another use of the progressive construction, which may have contributed to its increased frequency, can be found in the examples below.

- (3) I'm lovin' it! (McDonald's slogan)
- (4) I'm loving every moment with you.
- (5) Who're you wanting to seduce? (DCPSE:DI-C01 #0211:1:A)

Stative verbs like *love* and *want* do occur in the progressive, although for many speakers the simple present is still the expected form. This usage is not new. Denison (1998: 146) records some examples from as early as 1803 and 1820, and an example with the verb *love* from 1917. Nevertheless, it would be fair to say that until quite recently utterances like (3) and (4) would have been less marked if they contained a verb in the simple present tense.

Mufwene (1984:36) presents a 'scale of stativity', from punctual ('least stative') verbs to the 'highest stative' verbs as shown in (6) below (simplified).

- (6) **High:** e.g. *contain, know, belong to, consist of, need, concern, matter, owe*
- Intermediate:** e.g. *love, hate, depend, want, intend, wish*
- Neutral:** e.g. *enjoy, wait, stay, stand, lie; revolve, turn, work, run, read, write; call, claim, speak, say*
- Punctual/low:** e.g. *kick, reach, crack, die, break, hit, etc.*

The reports on the progressive discussed in section 2 above suggest that the progressive is spreading up the scale in PDE; currently it is often found with stative verbs such as *love, wish* and *want*, and perhaps in the future we will see an increase in use with verbs such as *know, need*, etc. Interestingly, in DCPSE, there is an example of *know* in the progressive from the 1961 data.

- (7) We will compare a play written in the Restoration Period with something that happened in Elizabethan times and we assume that our students *are knowing* what we are talking about you see. (DCPSE:DL-A01#0512)

## 5 Conclusions

In this paper we have shown how the *Diachronic Corpus of Present-day Spoken English* can be used to track short-term changes in the use of the progressive. The dataset shows an increased use of the construction in recent times, possibly due to a wider range of uses. In order to investigate this it would be necessary to explore the data further to differentiate these uses.

We have additionally demonstrated why it is very important to consider the frequency of the progressive relative to the possibility of it being used in the first place. Many studies have considered progressive use in terms of normalised absolute frequencies, such as the M-coefficient, i.e., frequencies considered in proportion to the number of words. However, the opportunity to use

any linguistic construction, including the progressive, may not be constant between different time periods or genres. The danger is that we end up measuring two things at the same time – (i) *the opportunity* to use the progressive combined with (ii) *the decision* to use the progressive, once the opportunity has arisen. Since we are interested in whether people increasingly choose to use the progressive, we must measure usage relative to opportunity.

A big advantage of using a parsed corpus like DCPSE is that in many cases the detailed grammatical analysis makes it easier to identify the set of cases where the opportunity for an event to occur arises. Ideally, we would wish to count the set of *true alternates*, i.e. those cases where we can say that the speaker could have chosen to use the progressive, but did not. This could be done by checking every VP in the corpus, or by estimating the total number of true alternates by inspecting a random subsample. Both options would be time-consuming. What we have done instead is assume that by discounting known VP forms which cannot take the progressive (*pace* Smitterberg) we arrive at the set of contexts where the option of using the progressive arises, thus obtaining a clearer picture of the increase in the use of the progressive in English.

## Notes

\* This research was carried out as part of the project *The changing verb phrase in present-day British English* funded by the UK Arts and Humanities Research Council (grant number AH/E006299/1).

1 ARCHER consists of written British and American English across a range of genres dating between 1650 and 1990. It was first constructed by Douglas Biber and Edward Finegan in the 1990s, and the latest release, ARCHER 3.1, was completed in 2006.

2 The upper line in Figure 1 shows absolute frequencies rather than M-coefficients, but Kranich asserts that, aside from the first half-century (1650-1700), the number of words per half-century do not differ greatly.

3 Press reportage consists of 44 texts, editorials 27 texts, and reviews 17 texts.

4 Mair and Leech (2006) claim that this trend is statistically significant. It is certainly large, but the data presented does not allow us to meaningfully draw this conclusion. Significance must be considered in relation to a choice between alternate forms. See section 3 for discussion.

5 The LLC is the spoken part of the *Survey of English Usage Corpus*, founded by Randolph Quirk in 1959. It contains 510,576 words of 1960s spoken English, is prosodically annotated, and has been used — and continues to be used — by many scholars for their research.

6 ICE-GB is composed of both spoken and written material from the 1990s. It contains textual markup, and is fully grammatically annotated.

7 Gloss (features are in italics): PU=parse unit, CL=clause, *main*=main, *intr*=intransitive; *ingp*=*ing* participle; SU=subject, NP=noun phrase, NPHD=NP head, PRON=pronoun, *pers*=personal, *plu*=plural; VB=verbal, VP=verb phrase, OP=operator; AUX=auxiliary; MVB=main verb, V=verb, *prog*=progressive; *encl*=enclitic; *pres*=present; A=adjunct; AVP=adverb phrase; AVHD=adverb phrase head; ADV=adverb; *ge*=general.

8 While the grammar that underlies the ICE-GB parsing (Quirk *et al.* 1985) conceives of Verb Phrases as only containing verbs (see Figure 2), in the proposed research the focus will be on the ‘extended VP’, i.e. a verb+dependents.

9 For more details on ICE-GB, DCPSE and ICECUP, see Aarts *et al.* (1998), Nelson *et al.* (2002), as well as [www.ucl.ac.uk/english-usage/projects/dcpse/research.htm](http://www.ucl.ac.uk/english-usage/projects/dcpse/research.htm).

10 Smitherberg (2005:47) also excludes non-finite VPs (progressive and non-progressive) on the grounds that in terms of complementation the choice seems to be between a *to*-infinitive and present participle (e.g. *She continued to read* vs. *She continued reading*) rather than a progressive and a non-progressive *to*-infinitive (e.g. *She continued to be reading* vs *She continued to read*). It seems to us that this depends on the verb in the super-ordinate clause; in PDE with a verb like *pretend* for example, the choice does seem to be between progressive and non-progressive (e.g. *She pretended to be reading* vs. *She pretended to read*). In order to exclude non-finite verb phrases, each example would have to be checked. This was beyond the scope of this paper.

11 Strictly speaking, we should add, *for language data consistent with the sampling of the corpus*. One possibility is that the way that samples were collected by Randolph Quirk *et al.* differs from the methodology of Sidney Greenbaum's team in the 1990s, and this explains the result.

12 The simple Binomial confidence interval for a probability (or percentage) is calculated by the following formula.

$$e = z_{crit} \sqrt{p(1-p)/N}$$

where  $z_{crit}$  is the critical value of  $z$  for a given confidence level,  $p$  is the probability of the event occurring (in this case, that the VP is progressive) and  $N$  is the total number of cases (i.e. applicable VPs).  $(1 - p)$  is the probability that the VP is *not* progressive. Note that for a 95% confidence interval,  $z_{crit}$  is approximately 1.96.

13 This scatter is limited (Pearson's  $r^2$ , fitting to a power law, is approximately ~95%). There are a number of sources of variance. Our samples are relatively small, the numbers of texts used in any given year are limited, and in DCPSE annual samples are not consistently balanced. Note that these sampling issues, while important to bear in mind, have not proved to be a barrier to obtaining this corpus-wide trend.

## References

- Aarts, Bas, María José López-Couso and Belén Méndez-Naya (forthcoming) “Late Modern English Syntax,” Volume 2 of *Historical Linguistics of English*, ed. by Brinton, Laurel, Mouton de Gruyter, Berlin and New York.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan (1999) *The Longman Grammar of Spoken and Written English*. Longman, London.
- Denison, David (1998) “Syntax,” *The Cambridge History of the English Language. IV: 1776-1997*, ed. by Romaine, Suzanne, 92-329, Cambridge University Press, Cambridge.
- Hundt, Marianne (2004) “Animacy, Agentivity, and the Spread of the Progressive in Modern English,” *English Language and Linguistics* 8.1, 47–69.
- Kranich, Svenja (2008) “The Progressive in Modern English. A Corpus-based Study of Grammaticalization and Related Changes,” Unpublished Doctoral Dissertation, Freie Universität Berlin.
- Labov, William (1972) *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.
- Leech, Geoffrey (2000) “Diachronic Linguistics across a Generation Gap: from the 1960s to the 1990s”, Paper read at the symposium Grammar and Lexis. University College London Institute of English Studies.
- Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith (2009) *Change in Contemporary English: A Grammatical Study*, Cambridge University Press, Cambridge.
- Mair, Christian (1995) “Changing Patterns of Complementation and Concomitant Grammaticalisation of the Verb *Help* in Present-day English,” *The Verb in Contemporary English*, ed. by Aarts, Bas, and Charles F. Meyer, 258-272, Cambridge University Press, Cambridge.
- Mair, Christian (1997) “Parallel Corpora: a Real-time Approach to the Study of Language Change in Progress,” *Corpus-based studies in English, Papers from the Seventeenth International Conference on English Language Research on Computerized Corpora*. ed. by Ljung, Magnus, 195-209, Rodopi, Amsterdam.
- Mair, Christian (2006) *Twentieth-Century English. History, Variation and Standardization*, Cambridge University Press, Cambridge.
- Mair, Christian and Marianne Hundt (1995) “Why is the Progressive Becoming More Frequent in English? A Corpus-based Investigation of Language Change in Progress,” *Zeitschrift für Anglistik und Amerikanistik* 43.2. 111-122.
- Mair, Christian and Marianne Hundt (1997) “The Corpus-based Approach to Language Change in Progress,” *Anglistentag 1996*, ed. by Böker, Uwe, and Hans Sauer, 71-82, Dresden.

- Mair, Christian and Geoffrey Leech (2006) "Current Changes in English Syntax," *The Handbook of English linguistics*, ed. by Aarts, Bas, and April McMahon, 318-342, Blackwell Publishers, Malden MA.
- Mossé, Ferdinand, (1938). *Histoire de la forme périphrastique être + participe présent en germanique*. 2 vols, Klincksieck, Paris.
- Mufwene, Salikoko (1984) "Stativity and the progressive," Indiana Linguistics Club, Bloomington, Indiana.
- Nehls, Dietrich (1988) "On the Development of the Grammatical Category of Verbal Aspect in English." *Essays on the English Language and Applied Linguistics. On the Occasion of Gerhard Nickel's 60<sup>th</sup> Birthday*, ed. by Klegraf, Josef and Dietrich Nehls, 173-198, Julius Groos Verlag, Heidelberg.
- Nesselhauf, Nadja (2007) "The Spread of the Progressive and its 'Future' Use," *English Language and Linguistics* 11, 1, 191–207.
- Núñez-Pertejo, Paloma (2007) "Aspects of the Use of the Progressive in the Eighteenth Century," *On varying language and opposing creed. New Insights into Late Modern English*, ed. by Pérez-Guerra, Javier, Dolores González-Alvarez, Jorge L. Bueno-Alonso and Esperanza Rama-Martínez, 359–382, Peter Lang, Bern.
- Smith, Nicholas (2005) "A Corpus-based Investigation of Recent Change in the Use of the Progressive in British English," Unpublished Doctoral Dissertation, Lancaster University.
- Smith, Nicholas and Geoffrey Leech (2001) "Grammatical Change in Recent Written English, Based on the FLOB and LOB Corpora," Paper read at the ICAME conference. Louvain-la-Neuve, Belgium.
- Smitterberg, Erik (2000) "The Progressive Form and Genre Variation During the Nineteenth Century," *Generative theory and corpus studies. A dialogue from 10 ICEHL*, ed. by Bermudez-Otero, Ricardo, David Denison, Richard Hogg and C. McCully, 283-297, Mouton de Gruyter, Berlin/New York.
- Smitterberg, Erik (2005) *The Progressive in 19<sup>th</sup>-Century English: A Process of Integration*. (Language and Computers: Studies in Practical Linguistics 54.) Amsterdam: Rodopi.
- Wallis, Sean (2003) "Scientific Experiments in Parsed Corpora: an Overview," *Extending the Scope of Corpus-based Research: New Applications, New Challenges, Language and Computers* 48, ed. by Granger, Sylviane and Stephanie Petch-Tyson, 12-23, Rodopi, Amsterdam.
- Wright, Susan (1994) "The Mystery of the Modal Progressive," *Studies in Early Modern English*. ed. by Kastovsky, Dieter, 467-485, Mouton de Gruyter, Berlin/New York.

Wright, Susan (1995) "Subjectivity and Experiential Syntax," *Subjectivity and Subjectivisation: Linguistic Perspectives*, ed. by Stein, Dieter and Susan Wright, 151-172, Cambridge University Press.