

The present progressive in political and academic language in the 19th and 20th centuries: A corpus-based investigation

*Erik Smitterberg
Uppsala University*

*Sabine Reich and Angela Hahn
Chemnitz University of Technology*

1 Introduction

1.1 Aim and scope

The aim of the present paper is twofold. First, the use of the progressive will be investigated in two genres from two different corpora, one of which consists of 19th-century English and one of Present-Day English (an English-German translation corpus). Focusing on the present progressive¹, we will look at how the normalized frequency of the construction differs between the two genres – political and academic texts – and between the two corpora. We will also examine the amount and type of modification of the present progressive form by temporal adverbials. Both frequency and adverbial modification are connected to the grammaticalization of the aspectual functions of the progressive form, which according to Nehls (1988: 188) took place ‘around the middle of the 19th century’.² When the progressive became grammaticalized for the expression of ongoing action, its frequency should reasonably have increased, but it is also possible that the necessity of qualifying the progressive with temporal adverbials should have decreased. Moreover, a few remarks on the role of temporal adverbials in German translations of the progressive will be made. Furthermore, we will take up some qualitative issues concerning the development of certain functions of the progressive form, such as emotional connotations and the ‘old’ and ‘new’ passive progressive, as in *The house is building* vs *The house is being built*.

Secondly, we wish to discuss features that affect corpus comparisons. When a study is based on two corpora with different sampling frames etc, to what

extent can the corpora be considered compatible, and what features affect their compatibility? If there are diachronic differences in results between the corpora, should they be attributed to actual linguistic development or to differences between the corpora used? In a diachronic study such as the present, the issue of genre development must also be taken into account: even in cases where the definition of a genre is roughly identical for both corpora, the genre itself may have changed over time.

1.2 Material

The material for the study comes from two different sources: the CONCE corpus and the English-German Translation Corpus. As part of the project ‘Style and Variation in Time and Space’ at Chemnitz University of Technology and Uppsala University³, aiming at enhancing existing corpus research methods, we have chosen corpora which have not been compiled with mutual compatibility in mind (cf the LOB/FLOB and Brown/Frown corpora). We hope this will give us opportunities to discuss what is a common problem for researchers who wish to investigate a linguistic feature in diachrony, ie the need to draw for data on several differently compiled corpora.

CONCE consists of roughly one million words of 19th-century British English, and is divided into seven genres (Debates, Drama, Fiction, History, Letters, Science, and Trials); of these, Debates and Science were selected for the present study. The Debates genre consists of recorded debates from the Houses of Parliament; Science consists of academic writing on topics pertaining to the natural and social sciences. CONCE is also divided into three subperiods (1800–1830, 1850–1870, and 1870–1900). However, since the genres selected for the present paper yielded comparatively few instances of present progressive forms, and since the diachronic dimension will be brought out by a comparison between the two corpora, the three subperiods were conflated for the present study. Moreover, only part of the Debates genre was used, since all of the first period and parts of the second consist of indirect speech and contain no present progressives. After these adjustments, the Debates genre yielded 14 present progressive forms and the Science genre 51.⁴

The English-German translation corpus (EGT) at Chemnitz is intended to represent present-day English and German; therefore most of the texts selected for the corpus were written within the last ten years. All English texts had been professionally translated into German. The majority of texts in this corpus are samples of written English. However, a number of scripts of speeches have been included, which may be classified as ‘written to be spoken’. The Chemnitz

English-German translation corpus comprises a variety of genres, an overview of which is given in Table 1:

Table 1: Descriptions of the genres in the English-German Translation Corpus

Genre	Description
Academic texts	Extracts from longer scientific and humanities texts
European Community texts	Information and policy documents
DOC Publications of the British Embassy, Bonn	Speeches, statements, debates and articles in the area of politics
Literature	Extracts from modern novels, detective novels, historical fiction and short stories
Tourist brochures	From various sources

The corpus currently comprises approximately one million words and is being expanded.

From this variety of genres we had to decide on some texts which are roughly comparable with those in CONCE. Moreover, if possible, we wanted to be able to compare speech-based and written language. For the present study we have therefore created two subcorpora of EGT which we will concentrate on. 18 speeches and one extract from a statement made during a parliamentary debate have been selected from the subcategory DOC – texts which may be classified as ‘written to be spoken’. This subcorpus, which will be called ‘Speeches and debates’, comprises 48,869 words. This provides us with a rough equivalent of the genre ‘Debates’ in CONCE. From the subcategory ‘Academic texts’, eight extracts from longer natural and social science texts have been selected, which amount to 166,429 words.

Table 2 summarizes the characteristics of each subcorpus, and it clearly shows that, even though the subcorpora for the 19th and 20th centuries are very similar, they are still not quite the same:

Table 2: The subcorpora selected from CONCE and EGT

Subcorpus	Genre characteristics
CONCE: Debates	Dialogue, speech written down ⁵ , formal
EGT: Speeches and debates	Monologue, written to be spoken, formal
CONCE: Science	Written to be read, formal
EGT: Natural and social science texts	Written to be read, formal

In what follows, in the interest of simplicity we will refer to Debates in CONCE and Speeches and debates in EGT taken together as ‘political language’; Science in CONCE and Natural and social science texts in EGT taken together will be referred to as ‘academic language’.

2 Results

2.1 Normalized frequencies

The commonest normalization procedure used with the raw frequency of the progressive form is the so-called M-coefficient (M after Mossé), where the raw frequency is normalized to a text length of 100,000 words. The results of applying this formula to the material can be found in Table 3:

Table 3: Normalized frequencies for present progressive forms across corpus and genre

	Political language			Academic language		
	Present progressives	Words	M-coefficient	Present progressives	Words	M-coefficient
CONCE	14	25,077	55.8	51	100,319	50.8
EGT	123	48,869	251.7	105	166,429	63.1

As can be seen in Table 1, in the 19th century the present progressive form is slightly more frequent in political language (Debates in CONCE) than in academic language (Science in CONCE). However, the low raw frequencies, especially for Debates in CONCE, make it impossible to draw any definitive conclusions. Moreover, it must also be borne in mind that the data for Debates were drawn from the second half of the 19th century only, whereas Science in CONCE also includes data from period 1 (1800–1830): as the progressive increased considerably in frequency over the 19th century, though with a great deal of variation across genres (see eg Strang 1982; Arnaud 1998; Smitterberg forthcoming), this discrepancy may also affect the results.

As far as the 20th-century data are concerned, the difference between political language (Speeches and debates in EGT) and academic language (Natural and social science texts in EGT) can be clearly seen. Our results are roughly comparable to what Hundt and Mair (forthcoming) found when investigating the frequency of the progressive in journalistic and academic prose in the LOB, FLOB, Brown, and Frown corpora, which suggests that, in speech-based and colloquialized written texts, the (present) progressive is far more frequent than in non-colloquialized written genres, such as academic texts. On a diachronic level, there seems to be a slight increase in the frequency of the present progressive in academic language, and a marked increase from the 19th to the 20th century in political language. Thus, the data basically confirm what has been observed about the increased frequency of the progressive in the literature.

2.2 Modification by temporal adverbials

The modification of the progressive form by temporal adverbials has been discussed by eg Crystal (1966), dealing with the verb phrase in general, Scheffer (1975), and Freckmann (1995)⁶, focusing on past progressives. However, opinions differ as to when an adverbial should be seen as modifying a progressive form. Freckmann (1995: 258) claims that, whereas Scheffer and Crystal automatically include any adverbial ‘that co-occurs with progressives within sentence limits’, he has checked his instances manually to make sure that the adverbials have scope over the progressives. On the other hand, Freckmann includes adverbials that occur in clauses (or, in speech, tone units) other than those containing a progressive form, provided that the adverbial is considered to modify the progressive. In the present study, the adverbial had both to occur in the same sentence as a progressive form and modify that progressive form in order to be included in the counts. As Scheffer (1975) provides separate figures for present progressives, a summary of his investigations will be given below.

Scheffer (1975: 184f) found 268 progressives in *Gregory's Dialogues*, 130 of which were modified by a temporal adverbial, yielding a modification percentage of 48.6; of the 34 present progressives, 13 were modified by temporal adverbials, yielding a percentage of 38.2.⁷ In a corpus consisting of ca 375,000 words of modern fiction, 523 of 2,468 progressive forms were modified by temporal adverbials, yielding a percentage of 21.2; of the 893 present progressives in the corpus, 136 had adverbial modification, yielding a percentage of 15.2 (Scheffer 1975: 54f). In a corpus of radio commentaries, finally, Scheffer found that 67 out of 167 progressive forms were modified by temporal adverbials, yielding a percentage of 40.1; of the 145 present progressives 58 were modified, yielding a percentage of 40.0 (Scheffer 1975: 121ff). Table 4 summarizes Scheffer's results:

Table 4: Modification of the progressive by temporal adverbials in Scheffer's (1975) corpora

	All progressives		Present progressives	
	Progressives	Adverbial modification	Progressives	Adverbial modification
<i>Gregory's Dialogues</i>	268	130 (48.6%)	34	13 (38.2%)
Modern fiction	2,468	523 (21.2%)	893	136 (15.2%)
Radio commentaries	167	67 (40.1%)	145	58 (40.0%)

It thus seems clear that adverbial modification of the progressive varies with both time and genre. If one extrapolates from Scheffer's figures, a possible hypothesis might be that older and/or speech-based texts favour adverbial modification of the progressive. The question therefore arises whether the figures from the present study will lead us to the same conclusion. The extent to which the present progressives – and the German translations of the present progressives in EGT – are modified by temporal adverbials in the present study is shown in Table 5.

Table 5: Modification of the present progressive form by temporal adverbials in political and academic language

	Political language		Academic language	
	Progressives	Adverbial modification	Progressives	Adverbial modification
CONCE	14	1 (7.1%)	51	13 (25.5%)
EGT (English)	123	43 (35.0%)	105	21 (20%)
EGT (German)	123 translations of the English verb into German (mostly present simple)	58 (47.2%)	105 translations of the English verb into German (mostly present simple)	21 (20%) ⁸

The raw frequencies for the 19th century are too small for any conclusions to be drawn from a cross-genre perspective. The 20th-century figures are probably high enough to allow some first tentative statements. The relative frequencies are in principle similar to Scheffer's figures; in particular, adverbial modification is more frequent in speeches/debates, ie written-to-be-spoken texts, than in the science section, a formal written genre. Hence, it seems to be the case that, as predicted, in genres close to spoken English, there is still a greater need to refer to the present moment or to make some other qualifications of time reference, duration or frequency (see also below).

From a diachronic perspective, it is at first glance more difficult to make any conclusive statements, because the figures appear to be contradictory. For the science subcorpora, we can observe a slight decrease of temporal adverbial modification, as expected. Hence, our results support the tendency which can already be detected in Scheffer's data: the modification of the progressive by temporal adverbials decreases with time. As our data show, this is not only true of literary language (Scheffer), but also of the language of science. In the speeches/debates section, however, the absolute and relative frequencies for the 19th and 20th centuries seem to suggest a contradictory result: adverbial modification of the present progressive has increased. However, for political language the scarcity of examples for the 19th century makes it impossible to draw conclusions, and renders statistical tests for significance unreliable, and for academic language, the differences between the figures are not statistically significant. The use of the chi-square test for statistical significance to test whether the variation might be owing to chance revealed that the difference was not significant even at the 90 per cent confidence interval.⁹

Looking at the adverbial modification in the German translations, we get yet another perspective on temporal adverbial modification. In German, adverbials play a far greater role for the question of how the verbal action is viewed in time. German – unlike English – does not have any linguistic means to express the concept of the present progressive as part of the verb phrase. Besides the various constructions in German which help to make the meaning of the present progressive explicit (Hahn, Reich and Schmied, in press), adverbial modification seems to be the most common option to emphasize the reference to the present time and thus express the meaning of an on-going process. Keeping this in mind, we can explain the increase of adverbials compared to English originals; in many cases the translator found it necessary to add an adverbial so that the meaning of the English present progressive is preserved in the German translation.

The most obvious option is to add the adverbs *jetzt* (now) or *zur Zeit* (at the moment), as in [1] and [2]:

- [1] That is what **is going on** in Bosnia and it must stop.
Doch gerade das **geschieht jetzt** in Bosnien, und das muß aufhören.
(doc9)
- [2] We **are inventing** something quite new and unique, and debate is therefore essential.
Wir **erfinden zur Zeit** etwas ganz Neues und Einmaliges, und daher sind Debatten unabdingbar.
(doc19)

Other adverbials have been chosen to express the reference to the present time, as example [3] shows (note that there is no time adverbial in the English original):

- [3] The interests succeeded the passions – as **is beginning to emerge from the researches of scholars** – as a means of disciplining and rendering them manageable and intelligible.
(pocb_e1.doc 55)

Autoren haben in jüngster Zeit gezeigt, daß man sein Interesse auf die Begierden richtete, um sie zu disziplinieren und einsehbar zu machen.
(pocb_2.doc)

(Autoren haben in jüngster Zeit gezeigt = ‘scholars have shown lately’)

Scheffer, following Crystal's (1966) classification, also groups the adverbials into three groups based on whether they denote frequency of occurrence, indicate restricted duration, or give a time-reference (Scheffer 1975: 50ff). Scheffer found that adverbials giving a time-reference are by far the commonest type in all three corpora. Our figures are too low to allow any definite conclusions, but one tendency clearly emerges: adverbials referring to the present time are the largest group among our temporal adverbials. Tables 6a and 6b give an overview of the variety of temporal adverbials attested in our corpora, ordered in descending frequency:

Table 6a: Temporal adverbials modifying the present progressive – political language

CONCE (Debates)		EGT (Speeches and debates)	
Adverbial	Frequency	Adverbial	Frequency
<i>constantly</i>	1	<i>now</i>	14
		<i>already</i>	8
		<i>still</i>	5
		<i>today</i>	3
		<i>also</i>	2
		<i>when-clause</i>	2
		<i>at this time</i>	1
		<i>at the same time</i>	1
		<i>this week</i>	1
		<i>while-clause</i>	1
		<i>year on year</i>	1
		<i>day by day</i>	1
		<i>as-clause</i>	1
		<i>again on 6 September</i>	1
		<i>for the first time in 25 years</i>	1

Table 6b: Temporal adverbials modifying the present progressive – academic language

CONCE (Science)		EGT (Natural and social science texts)	
Temporal adverbial	Frequency	Temporal adverbial	Frequency
<i>constantly</i>	4	<i>still</i>	5
<i>now</i>	2	<i>increasingly</i>	2
<i>still</i>	2	<i>actually</i>	2
<i>annually</i>	1	<i>now</i>	1
<i>daily</i>	1	<i>today</i>	1
<i>ever</i>	1	<i>at present</i>	1
<i>incessantly</i>	1	<i>currently</i>	1
<i>hourly</i>	1	<i>instantaneously</i>	1
<i>then</i>	1	<i>already</i>	1
<i>while-clause</i>	1	<i>forever</i>	1
		<i>sometimes</i>	1
		<i>usually</i>	1
		<i>constantly</i>	1
		<i>when-clause</i>	1
		<i>since then</i>	1

As these figures show, most adverbials serve to give a time reference. Apart from the ubiquitous *now*, speakers and writers seem to prefer a variety of options when referring to time. There are even examples of adverbials like *already* or *since then*, which are usually quoted in grammars as typically collocating with the present perfect (Lamprecht 1986: 260, 262). As example [4] shows, these adverbials are also used for modifying the present progressive:

- [4] I know that the business community **is already preparing** for new opportunities.
(doc6)

2.3 Discussion of special features

2.3.1 Voice

During the 19th century, the new type of progressive passive, as in *The house is being built*, largely ousted the older construction with active form and passive meaning, as in *The house is building* (Denison 1998: 148ff).¹⁰ However, among the present progressives in Debates and Science, there are actually no examples of the older construction; only the new, Present-Day English type is to be found. This is somewhat remarkable, as Visser (1973: 2016) claims that ‘the pattern of the type “the house is building” goes on occurring with great frequency in prose’ during the 19th century, and as there was plenty of negative prescriptive reaction to the new pattern (see Visser 1973: 2427f for exemplification). It is perhaps most surprising to see how clearly a formal written genre such as Science favours the type *The house is being built*, considering that Visser (1973: 2426f) claims that this type originated in spoken language and that ‘[e]ducated people who in the beginning used the new idiom may be conjectured as having done so only in familiar or unceremonious conversation with their intimate friends and the members of their family’, and that writers did not dare ‘to employ it in anything they knew was to be printed’. Moreover, Denison (1998: 153f) suggests that ‘[m]ost early examples tend to come from the pens of young people writing informally’, and hypothesizes that the type *The house is being built* may have been exploited by the social network of Robert Southey and Samuel Taylor Coleridge. However, all examples of the type *The house is being built* date from the latter half of the 19th century; it is therefore possible that the influence from the Southey/Coleridge circle had had time to contribute to the spread of the new progressive passive. The issue clearly merits further investigation, preferably combining a multi-genre approach, short-term diachronic methodology, and network analysis.

The distribution of active and passive progressive forms is shown in Table 7:

Table 7: Voice distribution of present progressive forms

	Political language		Academic language	
	Active	Passive	Active	Passive
CONCE	13 (92.9%)	1 (7.1%)	44 (86.3%)	7 (13.7%)
EGT	113 (91.9%)	10 (8.1%)	94 (89.5%)	11 (10.5%)

Even though the figures are very low, they suggest that the relative proportion of passive constructions among the present progressive has not changed. This gives further confirmation for the assumption that the present progressive passive had already been established in the 19th century. Since the figures are relatively low, especially for Debates in CONCE, this can only be regarded as a first tendency, not as a definite conclusion. There are also hardly any differences between the genres. The similar relative frequencies of EGT and CONCE are rather unexpected, considering the differences between the (sub-)corpora.

2.3.2 Attitudinal functions

Many scholars have commented on the progressive form having other functions than the purely aspectual, such as emphasis, emotion, vividness, politeness etc (see eg Wright 1994; Mindt 1997; Rydén 1997), which all express some type of subjective attitude towards what is being communicated. There are some examples which illustrate non-aspectual meaning in EGT, ie ‘shades of meaning’ of the present progressive:

- [5] Political power (also partly defined earlier) derives from the usefulness of centralized, institutionalized, territorialized regulation of many aspects of social relations. I **am not defining** it in purely “functional” terms, in terms of judicial regulation backed by coercion. Such functions can be possessed by any power organization – ideological, economic, military, as well as states. I restrict it to regulations and coercion centrally administered and territorially bounded – that is, to state power.
(mann_e1.doc, 40)
- [6] I **am proposing** that the UN should help the OAU and African countries to share this expertise and establish agreed mechanisms for ‘preventive diplomacy’ in Africa.
(doc4)

If we analyse these examples correctly, [5] and [6] are examples of the progressive used for expressing some reserve (or politeness/tact in Mindt’s terms). However, [5] and [6] can also be seen as ‘interpretative progressives’ (see eg König 1980; Ljung 1980). In these constructions, the progressive forms ‘give the speaker’s interpretation of some behaviour that somebody is engaged in’ (Ljung 1980: 69), a behaviour which in [5] and [6] is not overtly expressed. Similar instances can be found in CONCE:

- [7] I do not think that this proposal can possibly do any injustice to the Irish voter. We **are not seeking** to disfranchise anyone. We simply say that if

we give them votes in England they shall not also maintain their representation in Ireland to the same extent.

(*Debates*, 1870–1900, IV.1203)

As König (1980: 281) claims that ‘the “interpretive” use of the [progressive form] is a more recent phenomenon in the history of English’, and that ‘[i]t seems to have been rare in previous centuries’, it is interesting to see that it occurs in CONCE. It can also be found in academic language, as in [8]:

- [8] In propounding the fact, that the fluctuations in the prices of bills on a country, where the currency is depreciated, are governed, not only by all those elements of value previously considered, but to a much greater degree by the premium on bullion as compared with the depreciated currency, we **are stating** the final result rather than the process by which this position is reached – the ultimate limit, rather than the gradual development.

(*Science*, Goschen, 1850–1870, 62–63)

The present progressive is also used to express a certain tentativeness, which is clearly expressed in the German translation by adding the modal *sollte* (‘should’):

- [9] But if one employs the paradigm of classical politics, rather than that of natural jurisprudence in interpreting this great revolution in the concept of property – this transformation of the relations between polis and oikos, and between polity and economy, in the words of Joseph Cropsey – one **is not discounting** the importance of ideas about property derived from natural jurisprudence

(poca_e1.doc 10)

Doch das Politische mit dem Ökonomischen Seite an Seite zu stellen, um im Sprachgebrauch von Joseph Cropsey zu bleiben, **sollte nicht bedeuten**, ihr Verhältnis unbedingt als ein antithetisches aufzufassen; vielmehr möchte ich dahingehend argumentieren, daß diese Gegenüberstellung Gegenstand fortwährender und unabschließbarer Debatten ist

(pocb_d2.doc)

(*sollte nicht bedeuten* = ‘should not mean’)

Example [9] also has an interpretative effect; as with examples [5] and [7], the identity of the two acts is denied, not confirmed. Examples [5] through [9] all belong to one of the syntactic groups where the interpretative progressive occurs

listed by König. Examples [5] through [7] belong to the category ‘coordination, parataxis’, [8] belongs to the category ‘*in* + participle + main clause’, and [9] belongs to the category ‘Conditionals’ (König 1980: 275f).

One of the types of non-aspectual meaning that are commented on most frequently in the literature is the collocation of the progressive with an adverbial of the ALWAYS type (eg *always, ever, constantly, all the time*), which is usually said to give the utterance a subjective/emotive flavour. This collocation is used once to express a negative attitude in the Debates genre in CONCE, though with the slightly more formal *constantly* preferred to *always*:

- [10] We constantly hear querulous complaints of the law from hon. Gentlemen opposite, who were sent to this House by the operation of this law, and who represent what are called liberal opinions in this House. These hon. Gentlemen, we find, **are constantly quarrelling** with the political arrangements which sent them into the House.
(*Debates*, 1850–1870, 112.1176–112.1177)

The same collocation occurs six times in the Science genre (here too other adverbs are preferred to *always*). However, the intended effect is presumably general emphasis rather than emotion, given the more objective nature of the texts; it is also possible that no special effect is intended in some cases. Two examples are [11] and [12]:

- [11] So in the Russian exchanges, owing to the enormous amount of paper money afloat, which is practically inconvertible, the most violent fluctuations **are constantly occurring**.
(*Science*, Goschen, 1850–1870, 65)
- [12] THE vital statistics of a population are those of a vast army marching rank behind rank, across the treacherous table-land of life. Some of its members drop out of sight at every step, and a new rank **is ever rising up** to take the place vacated by the rank that preceded it, and which has already moved on.
(*Science*, Galton, 1870–1900, 164)

Corresponding examples are extremely rare in the EGT corpus, despite the higher absolute frequency of present progressives. Interestingly though, there are no examples of the ALWAYS modification type in the ‘speeches’ section of EGT, although one might expect such a modification in a genre which could be dubbed ‘persuasive’. In the science section of EGT, however, there are a few examples which clearly correspond to 19th-century usage:

- [13] [...] these pictures make it much easier to see how the atmosphere is **forever shifting** energy toward the poles.
(burr_e1.doc 90)
- [14] Sometimes we can be fairly clear about what parts of the content they do not want – e.g. in the case of the Irish people, the universal use of the Gaelic language – but such silent selective referenda are rarely possible. We **are constantly running** the risk of giving the people marks in terms of a syllabus they have not studied and an examination they are not taking.
(hobs_e1.doc 87)

Again, the progressive seems to be used for emphasis mainly.

3 Conclusion

The investigation of one linguistic feature, the progressive form, carried out in the present study indicates that it may be possible to use two corpora with different sampling frames for diachronic and cross-genre comparisons, provided that due consideration is given to the limitations of such an undertaking.

Our results have confirmed the view that the (present) progressive has become increasingly frequent during the 19th and 20th centuries. The present study also corroborates the frequent suggestion that the progressive is more common in speech-based than in written genres, and that this difference appears to increase over time. As concerns the passive voice, our corpus data suggest that the present progressive passive of the type *The house is being built* was more firmly established in the 19th century compared to the type *The house was building* than some scholars (cf Visser 1973: 2016f) appear to have thought. Our results indicate that there has been little diachronic variation as concerns active/passive voice distribution over the 19th and 20th centuries in the genres investigated. Attitudinal functions of the progressive are not very frequent in either corpus, but still a number of examples where the progressive has a polite or interpretative effect could be found both for the 19th and 20th centuries.

As far as adverbial modification of the present progressive is concerned, some interesting results emerge:

1. In both corpora adverbial modification is clearly attested, but no clear statistically significant differences or similarities across time, corpora, or genres emerged. Our data suggest a tendency for decreasing temporal adverbial modification of the present progressive. We see a possible connection to the fact that the progressive became grammaticalized during the

19th century and its expansion in the grammatical system rendered temporal adverbial modification less necessary. In political language, however, there were too few examples from the 19th century for the figures to be tested for statistical significance. It seems sensible to argue that, in genres which focus on the present moment and thus usually belong to spoken English, there is still a greater need to specify temporal reference.

2. As expected, the proportion of adverbials in German translations of the English progressive is higher than the proportion in the original texts. German needs adverbials to express aspectual meaning, since there is no way of indicating aspect on the verb in German.
3. Adverbs expressing time references are the most frequent type of adverbial modification. Apart from the stereotypical *now*, there is a very wide variety of adverbial modification, ranging from phrases like *at this time* to *when-* or *while*-clauses.

Despite these promising results, we would like to point out that our data must be approached with caution. First, the subcorpora used are relatively small, thus yielding a low number of constructions, and this problem is aggravated by the fact that the progressive is itself a low-frequency construction. Secondly, although a number of adjustments and selections have been made for the purpose of this study, the subcorpora used here are still slightly different, and there are several ways in which such differences may manifest themselves in the results. Hence, as concerns frequency and modification by adverbials of the ALWAYS type, the diverging figures for ‘Debates – CONCE’ and ‘Speeches – EGT’ might be a result of diachronic developments, but they could also be due to genre differences such as the following:

1. The Debates section in CONCE consists of recorded debates; ie we can assume a fair amount of dialogic features. In the ‘Speeches & debates’ section of EGT, however, monologues predominate.
2. The development of democracy and mass media both separate the periods covered by CONCE and EGT. It is possible that extralinguistic developments such as these may have affected the linguistic make-up of political language in general.
3. The sampling universes differ somewhat between the two corpora. It could be hypothesized that, if modification of the progressive by ALWAYS-type adverbials is more common in speech than in writing, texts which are written to be spoken (such as the speeches which largely constitute the ‘Speeches’ section in EGT) will contain fewer instances of this collocation

than texts which consist of speech taken down (such as the parliamentary debates that constitute the Debates genre in CONCE).

The sections of academic language in CONCE and EGT, however, are probably more comparable to each other; but it should be borne in mind that a number of factors may cause comparability problems even for these subcorpora:

1. The relative proportions of texts from different scientific fields may have consequences for the distribution of linguistic features. Biber (1988: 171) found that 'academic prose texts can be quite different from one another and still be considered representative of their genre', and that this relative lack of genre-internal coherence was at least partly accounted for by the existence of well-defined subgenres such as the natural and social sciences.
2. Academic writing taken as a genre has also changed in diachrony, though perhaps in less obvious ways than political language. Görlach (1999: 150) states that scientific style 'changed from somewhat personal accounts to impersonal, objective description' during the 19th century. Moreover, Biber and Finegan (1997) found that science texts have become more informational (as opposed to involved) (1997: 266) and less narrative (1997: 271) in style over the period 1650–1990.

There are thus several possible causes for the linguistic variation attested in the subcorpora used in the present study. Variation may be caused by linguistic and extralinguistic features, and the extralinguistic features may be under the researcher's control (such as sampling frames and genre definitions) or outside the researcher's control (such as diachronic genre development). We consider our results an interesting starting point for further discussion, and we hope that similar cross-corpora studies will help to confirm the tendencies we have detected.

Notes

1. The term 'present progressive' is used in the present study to denote a progressive form containing a present-tense form of the verb BE, eg *She is playing* and *He is being used*, but excluding progressive forms where the finite verb is in the present tense but consists of a perfect or modal auxiliary, eg *She has been playing* and *He may be being used*. We have decided to focus on the present progressive because its relatively high frequency in our corpora promises to yield representative results.

2. By ‘grammaticalization’ in this context we mean the process by which the progressive became, by and large, obligatory to express ongoing action in English.
3. We hereby gratefully acknowledge financial support for the project from DAAD and Svenska Institutet.
4. For this study, in constructions where part of the progressive form is shared between several verb phrases, such as *He was singing and dancing*, each partial verb phrase has been counted as a separate progressive form: the sentence *He was singing and dancing* is thus regarded as containing two progressive forms. If only the first of these were counted, ie *was singing*, the count would be 50 for the Science genre.
5. It must be kept in mind, however, that the Debates genre in CONCE does not constitute a subcorpus of spoken English proper. First, some of the debates are recorded as indirect rather than direct speech, thus increasing narrative features. Second, one must bear in mind the editorial process 19th-century spoken texts went through when being transferred into written form.
6. Freckmann seems to concern himself almost exclusively with temporal adverbials co-occurring with the progressive. However, there is nothing in his paper that expressly states that this is the case, and in fact *intentionally* and *evidently*, which are clearly not temporal adverbials, are also used in examples (Freckmann 1995: 257).
7. Scheffer provides separate counts for progressive forms which are modified by a temporal adverbial in ‘removed contexts’, ie outside the clause or sentence where the progressive occurs. Since such adverbials were not counted for the present study, adverbials in such ‘removed contexts’ have not been included in figures based on Scheffer’s counts.
8. The figures are slightly misleading here: in fact, there are three English adverbials which have not been translated into German, but there are also three additional German adverbials not attested in the English originals.
9. In CONCE as a whole, however, the percentage of progressive forms modified by temporal adverbials decreases over the 19th century.
10. The non-progressive passive, as in *The house is built*, was also an alternative (Denison 1998: 151). However, since the present study focuses on the progressive form, influence of non-progressive forms will not be dealt with.

References

1 Primary sources

CONCE = A Corpus of Nineteenth-Century English, being compiled by Merja Kytö (Uppsala University) and Juhani Rudanko (University of Tampere).

EGT = The English-German Translation Corpus, being compiled by Josef Schmied, Hildegard Schäffler, Barbara Fink and Diana Hudson-Ettle (Chemnitz University of Technology).

2 Secondary sources

Arnaud, René. 1998. The development of the progressive in 19th century English: A quantitative survey. *Language Variation and Change* 10, 123–152.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas and Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In T. Nevalainen and L. Kahlas-Tarkka (eds) *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253–275. Mémoires de la Société Néophilologique de Helsinki, Tome LII. Helsinki: Société Néophilologique.

Crystal, David. 1966. Specification and English tenses. *Journal of Linguistics* 2:1, 1–34.

Denison, David. 1998. Syntax. In S. Romaine (ed) *The Cambridge history of the English language volume IV: 1776–1997*, 92–329. Cambridge: Cambridge University Press.

Freckmann, Norbert. 1995. The progressive and adverbial collocations: corpus evidence. In W. Riehle and H. Keiper (eds) *Anglistentag 1994 Graz: Proceedings*, 255–268. Proceedings of the Conference of the German Association of University Teachers of English 16. Tübingen: Niemeyer.

Hahn, Angela, Sabine Reich and Josef Schmied. In press. Aspect in the Chemnitz Internet Grammar. *Proceedings of the ICAME conference 1999*, Freiburg.

Hundt, Marianne and Christian Mair. Forthcoming. ‘Agile’ and ‘uptight’ genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics*.

König, Ekkehard. 1980. On the context-dependence of the progressive in English. In C. Rohrer (ed) *Time, tense, and quantifiers: Proceedings of the*

- Stuttgart conference on the logic of tense and quantification*, 269–291. Linguistische Arbeiten 83. Tübingen: Max Niemeyer Verlag.
- Lamprecht, Adolf. 1986. *Grammatik der englischen Sprache*. 9th ed. Berlin: Cornelsen/Volk und Wissen.
- Ljung, Magnus. 1980. *Reflections on the English progressive*. Gothenburg Studies in English 46. Stockholm: Gotab.
- Mindt, Dieter. 1997. Complementary distribution, gradience and overlap in corpora and in ELT: Analysing and teaching the progressive. In U. Fries, V. Müller and P. Scheider (eds) *From Ælfric to the New York Times: Studies in English corpus linguistics*, 227–237. Language and Computers: Studies in Practical Linguistics 19. Amsterdam and Atlanta: Rodopi.
- Nehls, Dietrich. 1988. On the development of the grammatical category of verbal aspect in English. In J. Klegraf and D. Nehls (eds) *Essays on the English language and applied linguistics on the occasion of Gerhard Nickel's 60th birthday*, 173–198. Studies in Descriptive Linguistics 18. Heidelberg: Julius Groos Verlag.
- Rydén, Mats. 1997. On the panchronic core meaning of the English progressive. In T. Nevalainen and L. Kahlas-Tarkka (eds) *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 419–429. Mémoires de la Société Néophilologique de Helsinki, Tome LII. Helsinki: Société Néophilologique.
- Scheffer, Johannes. 1975. *The progressive in English*. Amsterdam: North-Holland.
- Smutterberg, Erik. Forthcoming. The progressive form and genre variation during the 19th century. In R. Bermúdez-Otero, D. Denison, R. M. Hogg and C. B. McCully (eds) *Generative theory and corpus studies: A dialogue from IOICEHL*. Topics in English Linguistics 33. Berlin and New York: Mouton de Gruyter.
- Strang, Barbara M. H. 1982. Some aspects of the history of the *be + ing* construction. In J. Anderson (ed) *Language form and linguistic variation: Papers dedicated to Angus McIntosh*, 427–474. Current Issues in Linguistic Theory 15. Amsterdam: John Benjamins.
- Visser, Fredericus Theodorus. 1973. *An historical syntax of the English language* III:2. Leiden: Brill.
- Wright, Susan. 1994. The mystery of the modal progressive. In D. Kastovsky (ed) *Studies in Early Modern English*, 467–485. Topics in English Linguistics 13. Berlin and New York: Mouton de Gruyter.