

# Diachronic Stylistic Changes in British and American Varieties of 20th Century Written English Language

Sanja Štajner and Ruslan Mitkov

Research Group in Computational Linguistics  
University of Wolverhampton  
Stafford Street, Wolverhampton, WV1 1SB, UK  
{S.Stajner, R.Mitkov}@wlv.ac.uk

## Abstract

In this paper we present the results of a study investigating the diachronic changes of four stylistic features: average sentence length, Automated Readability Index, lexical density and lexical richness in 20th century written English language. All experiments were conducted on the largest existing diachronic corpora of British and American English – the Brown ‘family’ corpora, employing NLP techniques for automatic extraction of the features. Additionally, we compare the trends of changes between the two English varieties and make suggestions for future studies of diachronic language change.

## 1 Introduction

As time elapses, language changes and those changes are present at various levels of the language structure: vocabulary, phonology, morphology and syntax (Kroch, 2001). Most of the previous studies of language change in English tended to focus on the phonetic and lexical rather than the stylistic or syntactic changes. Furthermore, the vast majority of the early sociolinguistic and historical linguistic studies of language change did not provide textual evidence for their claims, i.e. they were not corpus-based or the used corpora were not large and representative enough. Bauer (1994) set higher methodological standards for diachronic studies as he was among the first who sought to support his statements with textual evidence (Mair and Leech, 2006).

In his study of authorship attribution, Holmes (1994) defines style of the text “as a set of measurable patterns which may be unique to an author”. In the context of language change, this definition could be slightly amended and the style defined as a set of measurable patterns which may be unique to a particular period of time. In an extensive overview of applications in stylochro-metric approaches in the last sixty years (Stamou, 2008)

it is hypothesised that changes in certain aspects of an author’s writing style ought to be detectable by using appropriate methods and stylistic markers. Inspired by this hypothesis, one of the main goals of our study is to investigate whether diachronic changes in certain aspects of the writing style used in a specific text genre can be detected by using the appropriate methods and stylistic markers (features).

Different authors, even from the same period of time, will exhibit various styles. Consequently, this will be epitomised by very heterogeneous results in each observed year and genre, making general changes in style not readily detectable. Therefore, our first goal is to establish a methodology which would be appropriate for the investigation of these types of diachronic changes. This would lead to more precise and statistically justified conclusions in the corpus-based studies of diachronic linguistics.

Our second goal is to explore diachronic changes of several well-known stylistic markers: average sentence length, Automated Readability Index, lexical density and lexical richness in two major English language varieties – British and American and to detect whether those changes were present in both of those varieties. As we are using the mutually comparable corpora of British and American English, we are able to compare the changes of these two language varieties and to examine whether they followed the same trends of stylistic changes.

We base our methodology on the largest publicly available diachronic corpora of the 20th century written English language – Brown ‘family’ corpora (Leech and Smith, 2005) and NLP tools provided by the state-of-the-art Connexor Machine Syntax parser<sup>1</sup>.

---

<sup>1</sup>[www.connexor.eu](http://www.connexor.eu)

## 2 Related Work

Various studies in the field of stylistic variation and change use both the historical and sociolinguistic approach. Typical examples of this can be seen in the early studies of author's and period styles (Gordon, 1966; Adolph, 1968; Bennett, 1971). A preponderance of subsequent studies allude to the textual dimensions and relations used in Biber (1988) and follow the works of Biber and Finegan (1986, 1988, 1989). Westin (2002) and Westin and Geisler (2002) explore stylistic changes in the 20th century in the Corpus of English Newspaper Editorials (CENE) across the five linguistic dimensions described in (Biber and Finegan, 1988). They employ the methodology based on a multi-dimensional framework presented in (Biber, 1985; 1988).

The emergence of the FLOB and Frown corpora in the 1990s, compared with the previous LOB and Brown corpora, offered new possibilities for diachronic studies of written English language in the 20th century across two major English varieties – American and British English. Mair and Hundt (1995), Mair (1997, 2002), Mair, Hundt, Leech and Smith (2002), Smith (2002, 2003a, 2003b), Leech (2003, 2004), Leech and Smith (2006) and Mair and Leech (2006) exploited the possibilities of the Brown 'family' corpora by investigating the trends of changes in various lexical, grammatical and syntactic features. All of these studies were conducted on manually corrected POS tagged corpora and used the log likelihood test as a measure of the statistical significance of the results.

To our best knowledge, there is no mention of any investigation of diachronic changes of the four stylistic features we used in our study nor studies which describe automatic extraction of the features from the raw text corpora by using NLP techniques. All methodologies used in the previous studies of language changes required a great amount of human annotation or manual corrections which are time-consuming and labour-intensive or they relied on the use of very specific language tools.

## 3 Corpora

The Brown 'family' corpora (Leech and Smith, 2005) consist of four mutually comparable corpora publicly available as part of the ICAME Cor-

pus Collection<sup>2</sup>. Two of these contain texts written in British English, published in 1961 and 1991, respectively:

- The Lancaster-Oslo/Bergen Corpus of British English (**LOB**)
- The Freiburg-LOB Corpus of British English (**FLOB**)

The other two contain texts written in American English, published in 1961 and 1992, respectively:

- The Brown University corpus of written American English (**Brown**)
- The Freiburg-Brown Corpus of American English (**Frown**)

All four corpora share the same sampling frame, as the LOB<sup>3</sup>, FLOB<sup>4</sup> and Frown<sup>5</sup> corpora were designed to closely match the structure of the Brown corpus<sup>6</sup> with the aim to provide an opportunity to compare diachronic changes within two major varieties of the written English language (Leech and Smith, 2005).

Each corpus is a one-million-word corpus, consisting of 500 texts of about 2000 running words each, selected at a random point in the original source. The sampling range covers 15 text genres, which can be grouped into four more generalised categories:

- **Press** (Press: Reportage; Press: Editorial; Press: Review)
- **General Prose** (Religion; Skills, Trades and Hobbies; Popular Lore; Belles Lettres, Biographies, Essays)
- **Learned** (Miscellaneous; Science)
- **Fiction** (General Fiction; Mystery and Detective Fiction; Science Fiction; Adventure and Western; Romance and Love Story; Humour)

In this study we separately investigate diachronic stylistic changes in each of the four main text categories.

<sup>2</sup><http://www.hit.uib.no/icame>

<sup>3</sup><http://khnt.aksis.uib.no/icame/manuals/lob/>

<sup>4</sup><http://khnt.aksis.uib.no/icame/manuals/flob/>

<sup>5</sup><http://khnt.aksis.uib.no/icame/manuals/frown/>

<sup>6</sup><http://khnt.aksis.uib.no/icame/manuals/brown/>

## 4 Methodology

Although all four corpora are available in their tagged versions with the annotated sentence boundaries, those boundaries are not consistent throughout them. The LOB and FLOB tagged versions were manually corrected and therefore 100% accurate, while the Brown and Frown corpora are not manually corrected. Furthermore, some inconsistencies regarding the sentence boundaries in the cases of direct speech and itemised sentences are present among different corpora. Therefore, we decided to use the raw text versions of the LOB, FLOB, Brown and Frown corpora (divided into original portions of 500 texts of approximately 2,000 words each) and parse them with the Connexor Machine Syntax parser in order to achieve more consistent sentence splitting, tokenisation and lemmatisation. The sentence boundaries identified by the parser offered a fairer comparison of the results among the corpora which was of primary importance for this study.

Connexor Machine Syntax parser is based on linguistic methods and its lexicon contains hundreds of thousands of base forms. For compilation of the lexicon various large corpora were used, among which the most common were news texts, bureaucratic documents and literature (Connexor Oy, 2006b). In cases when the word cannot be found in the lexicon, word class and base form are determined by using the heuristic methods (Connexor Oy, 2006b). POS accuracy of the Machine Syntax parser measured on Standard Written English (benchmark from the Maastricht Treaty) was 99.3% with no ambiguity (Connexor Oy, 2006b). The earlier research of Samuelsson and Voutilainen (1997) reported excellent accuracy of the software used as the base for the current version of the parser. The fact that the lexicon of the Machine parser was built by using similar text genres as those represented in the Brown 'family' corpora ensures a high accuracy of the analysis completed in this study.

In the following two sub-sections we list some of the specificities of the tokenisation and lemmatisation procedures used by the Machine parser. This is important for better understanding and interpreting the results provided in this study.

### 4.1 Tokenisation

The Connexor Machine parser has a specific treatment of the contracted negative form and 's.

Contracted negative form and its antecedent verb are treated as separate tokens. E.g. in the case of *isn't*, the verb and negation are treated as two separate tokens *is* and *not* and correspondingly as two separate base forms *be* and *not*.

Treatment of 's depends on its role in the sentence (Connexor Oy, 2006a). In the cases where 's represents a genitive form, e.g. "... *Roy's United Federal Party ...*" (LOB:A01), *Roy's* is treated as one token *Roy's* and corresponding base form is *roy*. In other cases where 's represents the contraction of the verb *to be* (*is*) or *to have* (*has*), e.g. "... *he's nice ...*" (LOB:P05), the personal pronoun and verb contraction are treated as two separate tokens – *he* and *is*. Accordingly, they are treated as two separate base forms – *he* and *be*.

### 4.2 Lemmatisation

In this sub-section, we describe the results of the lemmatisation process for the three word types – possessive pronouns, derived adverbs and EN and ING forms, as they differ between the current and the previous versions of the parser.

#### 4.2.1 Base forms of possessive pronouns and derived adverbs

The current version of the Machine parser assigns to possessive pronouns their own base forms (lemmas), e.g. the base form of the word *yours* in the current version of the parser is *yours*, while in the previous versions possessive pronouns were assigned the base forms of their corresponding personal pronouns, e.g. the base form of the word *yours* in the previous versions of the parser was *you* (Connexor Oy, 2006b).

Base forms of derived adverbs, e.g. *immediately*, *fundamentally*, *absolutely*, *directly* (LOB:A01) are assigned *immediately*, *fundamentally*, *absolutely*, *directly* as their base forms in the current version of the parser while in the previous versions they were assigned the following base forms: *immediate*, *fundamental*, *absolute*, *direct*.

#### 4.2.2 EN and ING forms

The current version of the parser assigns different base forms for EN and ING forms which represent the present or past participle of a verb than for those representing corresponding nouns or adjectives. Previous versions of the parser were assigning the same base forms for all four possible cases of EN and ING forms. For example, in the sentence:

“The Government decided to adjust the financing ...” (LOB:A01)

the assigned base form for the noun *financing* is *financing* while the older versions of the parser would assign *finance* as the base form for the same word *financing* in the given context.

Similarly, in the sentence:

“Sir Roy is violently opposed to Africans getting an elected majority in Northern Rhodesia ...” (LOB:A01)

the assigned base form for the adjective *elected* is *elected* while the older versions of the parser would assign *elect* as the base form for the same word *elected* in the given context.

In the above sentence, the words *opposed* and *getting* are assigned *oppose* and *get* as the base forms in all versions of the Machine parser, as they represent the past and present participle of the verbs *oppose* and *get*, respectively.

## 5 Experimental settings

We conducted two sets of experiments:

- Stylistic diachronic changes in British English in the period 1961–1991
- Stylistic diachronic changes in American English in the period 1961–1992

The first experiment was conducted on LOB (1961) and FLOB (1991) corpora and the second experiment on Brown (1961) and Frown (1992) corpora.

In both experiments we investigated the diachronic changes over the four main text categories: Press, General Prose, Learned and Fiction separately, as the preliminary results had shown different trends of changes among these four text categories.

As stylistic markers we used the following four features:

- Average sentence length (ASL)
- Automated Readability Index (ARI)
- Lexical density (LD)
- Lexical richness (LR)

**Average sentence length** was measured as the total number of words divided by the total number of sentences for each text (eq.1), using the sentence and word boundaries returned by the parser. Words containing only punctuation marks were not counted.

$$ASL = \frac{total\_number\_of\_words}{total\_number\_of\_sentences} \quad (1)$$

**Automated Readability Index** (Senter and Smith, 1967; Smith and Kincaid, 1970) belongs to the formative era of readability studies and was listed among eleven most commonly used readability formulas (McCallum and Peterson, 1982) of that time. It is calculated using the following formula:

$$ARI = 4.71 \frac{c}{w} + 0.5 \frac{w}{s} - 21.43 \quad (2)$$

where  $c$ ,  $w$  and  $s$  represent, respectively, total number of characters, words and sentences in the text. The result of the Automated Readability Index gives the US grade level necessary to understand the given text.

**Lexical density** is computed as the number of unique word types (tokens) divided by the total number of tokens in the text (eq.3).

$$LD = \frac{number\_of\_unique\_tokens}{total\_number\_of\_tokens} \quad (3)$$

Low lexical density indicates many repetitions of the same words throughout the text, while high lexical density suggests a use of a wider range of vocabulary. This feature has been used as a stylistic marker in (Ule, 1982) and for dating works in (Smith and Kelly, 2002).

**Lexical richness** is computed as the number of unique lemmas divided by the total number of tokens in the text (eq.4).

$$LR = \frac{number\_of\_unique\_lemmas}{total\_number\_of\_tokens} \quad (4)$$

The use of lexical richness separately from lexical density was proposed by Corpas Pastor et al. (2008) who argued that “lexical density is not indicative of the vocabulary variety of an author as it counts morphological variants of the same word as different word types”. Following this argument, we make a distinction between lexical density and lexical richness and investigate both features separately.

## 6 Results and Discussion

The results of the first and second experiment are given separately (sub-sections 6.1 and 6.2, respectively). The results of each experiment are divided into two tables, where the first table contains the results of the investigation of diachronic changes of ASL and ARI, and the second table contains results of the investigation of diachronic changes of LD and LR. Each feature contains two columns – ‘change’ and ‘p’.

Column ‘change’ represents the relative change of the feature over the period 1961–1991/2, measured as a percentage of the starting value in 1961. The sign before the value signifies the direction of the change; ‘+’ corresponds to an increase and ‘–’ to a decrease. Both the starting and ending values in years 1961 and 1991/2, respectively, were previously calculated as an arithmetic mean of the feature values in all texts of the relevant text category (Press, Prose, Learned or Fiction) and corpus (LOB or FLOB in the sub-section 6.1 or Brown and Frown in the sub-section 6.2).

Column ‘p’ represents the p-value of the two-tailed t-test applied on the results obtained for each text separately (in the corresponding text category and corpus) and used as a measure of statistical significance. For the results with a p-value lower than the chosen critical value 0.05, we are more than 95% sure that they represent real diachronic changes rather than being a consequence of faulty sampling. Those results were considered as statistically significant and reliable enough to be used for making further hypotheses.

Given that the corpora used in the two experiments are mutually comparable, we are able to make a comparison of the trends of changes between British and American English in sub-section 6.3.

### 6.1 Diachronic changes in British English

Results of the first experiment – diachronic stylistic changes in British English over the period 1961–1991 are presented in Table 1 and Table 2.

Genre	ASL		ARI	
	change	p	change	p
Press	+4.39%	0.114	+ <b>6.43%</b>	<b>0.046</b>
Prose	+1.64%	0.490	+ <b>9.07%</b>	<b>0.002</b>
Learned	–5.05%	0.060	+3.07%	0.254
Fiction	–4.85%	0.184	–1.97%	0.726

Table 1: British English: ASL and ARI.

Automated Readability Index shows a statistically significant increase over the observed period 1961–1991 in the Press and Prose text categories (Table 1) which can be interpreted as a tendency to render texts in these categories in a difficult-to-read manner.

The results of the first experiment (Table 1) indicate that ASL did not change significantly in the period 1961–1991 in any of the four investigated text categories of British English.

Genre	LD		LR	
	change	p	change	p
Press	+ <b>7.43%</b>	<b>0.000</b>	+ <b>7.17%</b>	<b>0.000</b>
Prose	+ <b>3.90%</b>	<b>0.000</b>	+ <b>4.40%</b>	<b>0.000</b>
Learned	+2.35%	0.248	+1.76%	0.416
Fiction	+ <b>3.92%</b>	<b>0.008</b>	+ <b>4.28%</b>	<b>0.012</b>

Table 2: British English: LD and LR.

On the basis of the data analysed (Table 2), we can draw the conclusion that the vocabulary was enriched in three text categories of British English – Press, Prose and Fiction, over the observed period 1961–1991. The strongest intensity of these changes can be noticed in the Press category.

### 6.2 Diachronic changes in American English

Results of the second experiment – diachronic stylistic changes in American English over the period 1961–1992 are presented in Table 3 and Table 4.

Genre	ASL		ARI	
	change	p	change	p
Press	– <b>4.90%</b>	<b>0.034</b>	–1.43%	0.598
Prose	–3.35%	0.164	+3.32%	0.242
Learned	– <b>10.49%</b>	<b>0.000</b>	–3.29%	0.278
Fiction	–7.37%	0.058	–9.92%	0.082

Table 3: American English: ASL and ARI.

The Press and Learned text categories manifested a decrease of ASL in the observed period 1961–1992 (Table 3). This could be interpreted as an example of colloquialisation – “a tendency for the written language gradually to acquire norms and characteristics associated with the spoken conversational language” (Leech, 2004), as it is known that shorter sentences are a characteristic of the spoken language.

The results in Table 3 indicate that ARI did not change significantly in the period 1961–1991 in any of the four investigated text categories of American English.

Lexical density and lexical richness demonstrated a statistically significant increase only in the Prose text category (Table 4).

Genre	LD		LR	
	change	p	change	p
Press	+2.03%	0.084	+0.71%	0.588
Prose	<b>+4.06%</b>	<b>0.000</b>	<b>+3.85%</b>	<b>0.004</b>
Learned	+3.89%	0.082	+3.90%	0.092
Fiction	-2.51%	0.106	-3.04%	0.088

Table 4: American English: LD and LR.

These results lead to the conclusion that the vocabulary used in the Prose category of American English was enriched over the observed period 1961–1992.

### 6.3 British vs. American diachronic changes

From the results presented in the above four tables we are able to make a comparison of the trends of diachronic changes between British and American English for the four investigated stylistic features and provide some general observations.

The central conclusion is that British and American English do not follow the same trends of stylistic diachronic changes in all genres and for all features. The most striking differences are in the cases of ASL and ARI. They demonstrated statistically significant changes in only one of the two investigated varieties of the English language. ASL had a statistically significant decrease in the Press and Learned categories of American English over the period 1961–1991 (Table 3), while at the same time did not manifest any statistically significant changes in any of the four text categories of British English (Table 1). ARI demonstrated a statistically significant increase in the Press and Prose categories of British English (Table 1) and no statistically significant changes in any of the four text categories of American English (Table 3).

LD and LR manifested a statistically significant increase in three text categories of British English – Press, Prose and Fiction (Table 2). In American English, this trend was followed only in the Press category, while the other two text categories – Prose and Fiction, did not exhibit any statistically significant changes of LD or LR (Table 4).

In order to better understand noticed diachronic changes and investigate possible influence between the two English language varieties, we conducted an additional experiment of synchronic comparison between British and American En-

glish in both starting and ending years – 1961 and 1991/2, for each genre and each feature separately. These results were consistent with the results of the previous two experiments – diachronic changes in British and diachronic changes in American English, thus supporting the hypotheses made in the previous two subsections.

Of most interest were the results obtained from the Press category (Table 5), as they suggested the presence of Americanisation – “the influence of north American habits of expression and behaviour on the UK (and other nations)” (Leech, 2004) in the observed period 1961–1991/2. They indicated that the influence between the British and American English written styles was more pronounced in the Press category than in the other three – Prose, Learned and Fiction.

Feature	1961		1991/2	
	Br.	Am.	Br.	Am.
ASL	20.63	21.50	<b>21.53*</b>	<b>20.45</b>
ARI	<b>11.34</b>	<b>12.08*</b>	12.07	11.91
LD	<b>36.37</b>	<b>37.65*</b>	39.08	38.41
LR	<b>32.61</b>	<b>33.83*</b>	34.95	34.07

Table 5: Synchronic comparison: Press category.

In Table 5, the value of the LD and LR calculated using equations 3 and 4 (Section 5) is multiplied by 100. Statistical significance of the differences between the feature values in British and American English is measured by the two-tailed t-test. The results significant at the 0.05 level are shown in bold. In those cases, the higher of the two values (Br. or Am.) is marked with an ‘\*’.

Statistically significant increases of the ARI, LD and LR in the Press category of British English during the period 1961–1991 (Table 1 and Table 2) together with the significantly higher values of these features in the same text category of American English in 1961 (Table 5) could be explained using the aforementioned Americanisation hypothesis.

## 7 Conclusions

On the basis of the data analysed in the previous section, we can draw the following conclusions about the trends of stylistic changes in the British and American varieties of English language over the period 1961–1991/2:

1) The Prose text category followed the same trend of enriching the vocabulary in both varieties of the English language.

2) Average sentence length had a statistically significant decrease in the Press and Learned text categories in American English and no statistically significant changes in any of the four text categories in British English.

3) Automated Readability Index had a statistically significant increase in the Press and Prose text categories in British English and no statistically significant changes in any of the four text categories in American English.

Additionally, the presented study allowed us to make several more general conclusions:

1) NLP tools can be successfully used in the studies of language change and make use of the raw text corpora customising it for the specific purposes thus saving a great amount of human effort for annotation or manual corrections.

2) Stylistic changes are present and are noticeable even after the 30 year time gap in various categories of the written language.

3) Different genres show different trends of stylistic changes over the same period of time. Therefore, it is of great importance to investigate them separately in order to obtain a better picture of the process of language change and the possible influences between different genres.

4) As different language varieties show different trends of stylistic changes inside the same text categories, no general conclusions about the trends of stylistic changes should be made before a detailed investigation in each of the language varieties. Furthermore, results of the separate investigations of stylistic changes among different language varieties enable a better understanding of the noticed trends and their possible mutual influence. However, it is important to ensure that the corpora of different language varieties are mutually comparable and thus any similarities or differences among their trends of change are not due to different sampling methods and text selection.

Finally, we demonstrated the possibility of using the Brown 'family' corpora and NLP techniques for the investigation of diachronic stylistic changes. It has created a path for many other stylistic features to be investigated using the same corpora and utilising the full potential of the current state-of-the-art NLP tools and techniques.

## Acknowledgements

This project was supported by the European Commission, Education & Training, Erasmus Mundus:

EMMC 2008-0083, Erasmus Mundus Masters in NLP & HLT.

## References

- Adolph Robert. 1966. *The Rise of Modern Prose Style*. Cambridge, Mass.: M.I.T. Press.
- Bauer Laurie. 1994. *Watching English change: and introduction to the study of linguistic change in standard English in the twentieth century*. London: Longman.
- Bennett R. James. 1971. *Prose Style: A Historical Approach through Studies*. San Francisco: Chandler.
- Biber Douglas. 1985. Investigating Macroscopic Textual Variation through Multifeature / Multidimensional Analyses. *Linguistics*, 23: 337–360.
- Biber Douglas and Finegan Edward. 1986. An Initial Typology of English Text Types. In: J. Aarts and W. Meijs, eds. *Corpus Linguistics H: New Studies in the Analysis and Exploitation of Computer Corpora*. Amsterdam, Rodopi, 19–46.
- Biber Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber Douglas and Finegan Edward. 1988. Drift in three English genres from the 18th to the 20th century: A multi-dimensional approach. In: M. Kyt, O. Ihalainen, and M. Rissanen, eds. *Corpus linguistics, hard and soft. Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora*: 83–101. Amsterdam: Rodopi.
- Biber Douglas and Finegan Edward. 1989. Drift and the evolution of English style: A history of three genres. *Language*, 65: 487–517.
- Connexor Oy. 2006a. *Machine Language Model*.
- Connexor Oy. 2006b. *Machine Language Analysers*.
- Corpas Pastor Gloria, Mitkov Ruslan, Afzal Navid and Pekar Viktor. 2008. Translation Universals: Do they exist? A corpus-based NLP study of convergence and simplification. In *Proceedings of the AMTA*. Waikiki, Hawaii.
- Gordon A. Ian. 1966. *The Movement of English Prose*. Bloomington: Indiana University Press.
- Holmes I. David. 1994. Authorship Attribution. *Computers and the Humanities*, 28(2): 87–106. Springer Netherlands.
- Kroch Anthony. 2001. Syntactic change. In: M. Baltin and C. Collins, eds. *The Handbook of Contemporary Syntactic Theory*. Malden, MA: Blackwells, 699–730.

- Leech Geoffrey. 2003. Modality on the move: the English modal auxiliaries 1961-1992. In: R. Facchinetti, M. Krug and F. Palmer, eds. *Modality in contemporary English*. Berlin/New York: Mouton de Gruyter, 223-240.
- Leech Geoffrey. 2004. Recent grammatical change in English: data, description, theory. In: K. Aijmer and B. Altenberg, eds. *Advances in Corpus Linguistics: Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23) Gteborg 22-26 May 2002*. Amsterdam: Rodopi, 61-81.
- Leech Geoffrey and Smith Nicholas. 2005. Extending the possibilities of corpus-based research on English in the twentieth century: a prequel to LOB and FLOB. *ICAME Journal*, 29: 83-98.
- Leech Geoffrey and Smith Nicholas. 2006. Recent grammatical change in written English 1961-1992: some preliminary findings of a comparison of American with British English. In: A. Renouf and A. Kehoe, eds. *The Changing Face of Corpus Linguistics*. Amsterdam: Rodopi, 186-204.
- Mair Christian and Hundt Marianne. 1995. Why is the progressive becoming more frequent in English? A corpus-based investigation of language change in progress. *Zeitschrift fr Anglistik und Amerikanistik*, 43: 111-122.
- Mair Christian. 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 linguistic modelling: a festschrift for Jacek Fisiak on his 60th birthday*. Berlin: Mouton de Gruyter, 1536-1543.
- Mair Christian. 2002. Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora. *English Language and Linguistics*, 6: 105-131.
- Mair Christian, Hundt Marianne, Leech Geoffrey and Smith Nicholas. 2002. Short term diachronic shifts in part-of-speech frequencies: a comparison of the tagged LOB and F-LOB corpora. *International Journal of Corpus Linguistics*, 7: 245-264.
- Mair Christian and Leech Geoffrey. 2006. Current change in English syntax. In: B. Aarts and A. MacMahon, eds. *The Handbook of English Linguistics*, Ch.14. Oxford: Blackwell.
- McCallum R. Douglas and Peterson L. James. 1982. Computer-based readability indexes. In *Proceedings of the ACM '82 Conference*: 44-48. New York, NY.
- Samuelsson Christer and Voutilainen Ato. 1997. Comparing a linguistic and a stochastic tagger. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, (ACL '98): 246-253. Association for Computational Linguistics, Stroudsburg, PA, USA.
- Senter, R. J. and Smith E. A. 1967. *Automated readability index*. Technical Report (AMRLTR-66-220). University of Cincinnati, Cincinnati: Ohio.
- Smith A. Edgar and Kincaid J. Peter. 1970. Derivation and Validation of the Automated Readability Index for Use with Technical Materials. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 12(5): 457-464.
- Smith A. Joseph and Kelly Colleen. 2002. Stylistic Constancy and Change Across Literary Corpora: Using Measures of Lexical Richness to Date Works. *Computers & the Humanities*, 36(4): 411-431.
- Smith Nicholas. 2002. Ever moving on? The progressive in recent British English. In: P. Peters, P. Collins and A. Smith, eds. *New frontiers of corpus research: papers from the twenty first International Conference on English Language Research on Computerized Corpora, Sydney 2000*. Amsterdam: Rodopi, 317-330.
- Smith Nicholas. 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*: 714-723. Lancaster University: UCREL Technical Papers.
- Smith Nicholas. 2003b. Changes in the modals and semi-modals of strong obligation and epistemic necessity in recent British English. In: R. Facchinetti, M. Krug and F. Palmer, eds. *Modality in contemporary English*. Berlin/New York: Mouton de Gruyter, 241-266.
- Stamou Constantina. 2008. Stylochronometry: Stylistic Development, Sequence of Composition, and Relative Dating. *Literary & Linguistic Computing*, 23(2): 181-199.
- Ule Louis. 1982. Recent progress in computer methods of authorship determination. *Association of Literary and Linguistic Computing Bulletin*, 10(3): 73-89.
- Westin Ingrid and Geisler Christer. 2002. A multi-dimensional study of diachronic variation in British newspaper editorials. *ICAME Journal*, 26: 133-152.
- Westin Ingrid. 2002. *Language Change in English Newspaper Editorials*. Amsterdam: Rodopi.