Style of Religious Texts in 20th Century

Sanja Štajner, Ruslan Mitkov

Research Institute in Information and Language Processing
University of Wolverhampton, UK
S.Stajner@wlv.ac.uk, R.Mitkov@wlv.ac.uk

Abstract

In this study, we present the results of the investigation of diachronic stylistic changes in 20th century religious texts in two major English language varieties – British and American. We examined a total of 146 stylistic features, divided into three main feature sets: (average sentence length, Automated readability index, lexical density and lexical richness), part-of-speech frequencies and stop-words frequencies. All features were extracted from the raw text version of the corpora, using the state-of-the-art NLP tools and techniques. The results reported significant changes of various stylistic features belonging to all three aforementioned groups in the case of British English (1961–1991) and various features from the second and third group in the case of American English (1961–1992). The comparison of diachronic changes between British and American English pointed out very different trends of stylistic changes in these two language varieties. Finally, the applied machine learning classification algorithms indicated the stop-words frequencies as the most important stylistic features for diachronic classification of religious texts in British English and made no preferences between the second and third group of features in diachronic classification in American English.

Keywords: stylometry, language change, text classification

1. Introduction

According to Holmes (1994), style of the text could be defined "as a set of measurable patterns which may be unique to an author". Štajner and Mitkov (2011) further amended this definition to the scope of the language change by defining style "as a set of measurable patterns which may be unique in a particular period of time" and tried to examine whether "certain aspects of the writing style used in a specific text genre can be detected by using the appropriate methods and stylistic markers". In that study, they investigated only four stylistic features over the four main text genres represented in the 'Brown family' of corpora – Press, Prose, Learned and Fiction. In this study, we followed their main ideas and methodology but focused only on the genre of religious texts. We made a more in depth analysis of stylistic changes by using more features (a total of 146 features) and applied machine learning techniques to discover which features underwent the most drastic changes and thus might be the most relevant for a diachronic classification of this text genre.

1.1. Corpora

The goal of this study was to investigate diachronic stylistic changes of 20th century religious texts in British and American English and then compare the trends of reported changes between these two language varieties. Therefore, we used the relevant part (genre D – Religion in Table 1) of the only publicly available corpora which fulfills both conditions of being diachronic and comparable in these two English language varieties – the 'Brown family' of corpora. The British part of the corpora consists of the following three corpora:

- The Lancaster1931 Corpus (BLOB),
- The Lancaster-Oslo/Bergen Corpus (LOB)
- The Freiburg-LOB Corpus of British English (FLOB).

These corpora contain texts published in 1931$\pm$3, 1961 and 1991, respectively. The American part of the 'Brown family' of corpora consists of two corpora:

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

These two corpora contain texts published in 1961 and 1992, respectively. Four of these corpora (LOB, FLOB, Brown and Frown) are publicly available as a part of the ICAME corpus collection[1] and they have been widely used across the linguistic community for various diachronic and synchronic studies as they are all mutually comparable (Leech and Smith, 2005). The fifth corpus (BLOB) is still not publicly available, although it has already been used in some diachronic studies, e.g. (Leech and Smith, 2009).

As the initial purpose of compiling the Brown corpus was to have a representative sample of 'standard' English language (Francis, 1965), the corpus has covered 15 different text genres, which could further be clustered into four main text categories – Press, Prose, Learned and Fiction (Table 1). The other four corpora which were compiled later (LOB, FLOB, Frown and BLOB) shared the same design and sampling method with the Brown corpus, thus making all five of them mutually comparable.

The genre which is relevant for this study – genre D (Religion), belongs to the broader Prose text category (Table 1). To the best of our knowledge, this genre has never been used in any diachronic study on its own, instead it was always included as a part of the Prose category, together with

[1] http://icame.uib.no/newcd.htm

Category	Code	Genre
PRESS	A	Press: Reportage
	B	Press: Editorial
	C	Press: Review
PROSE	D	Religion
	E	Skills, Trades and Hobbies
	F	Popular Lore
	G	Belles Lettres, Biographies, Essays
	H	Miscellaneous
LEARNED	J	Science
FICTION	K	General Fiction
	L	Mystery and Detective Fiction
	M	Science Fiction
	N	Adventure and Western
	P	Romance and Love Story
	R	Humour

Table 1: Structure of the corpora

the other four Prose genres (E–F). Although this genre contains only 17 texts of approximately 2,000 words each, the texts were chosen in the way that they cover different styles and authors of religious texts. For instance, in the Brown corpus, 7 of those texts were extracted from books, 6 from periodicals and 4 from tracts[2]. The full list of used texts and the authors for each of the four corpora could be found following the links given in their manuals[3]. Although the size of the corpora used in this study (approx. 34,000 words in each corpus) is small by the present standards of corpus-based research, it is still the only existing diachronic comparable corpora of religious texts. Therefore, we find the results presented in this paper relevant though we suggest that they should be considered only as preliminary results until a bigger comparable corpora of religious texts become available.

1.2. Features

In this study, we focused on genre D (Religion) of the four publicly available parts of the 'Brown family' of corpora and investigated diachronic stylistic changes in British (using the LOB and FLOB corpora) and American (using the Brown and Frown corpora) English. As the LOB and FLOB corpora cover the time span from 1961 to 1991, and the Brown and Frown corpora from 1961 to 1992, we were also able to compare these diachronic changes between the two language varieties in the same period 1961–1991/2. In both cases, we used three sets of stylistic features. The first set contains features previously used by Štajner and Mitkov (2011):

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

Average sentence length has been used as a feature for stylistic categorisation and authorship identification since 1851 (Holmes, 1998; Gamon, 2004). It is calculated as the total number of words divided by the total number of sentences in the given text (eq.1).

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

Automated Readability Index (Senter and Smith, 1967; Kincaid and Delionbach, 1973) is one of the many readability measures used to assess the complexity of the texts by giving the minimum US grade level necessary for its comprehension. McCallum and Peterson (1982) have listed it among eleven most commonly used readability formulas of that time, which was probably related to the fact that it is very easy to be computed automatically. Unlike the other readability indexes which usually require the number of syllables in text (difficult to compute automatically with a high precision), ARI only requires the number of characters (c), words (w) and sentences (s) in the given text (eq.2).

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

Lexical density has already been in use as a stylistic marker in, e.g. (Ule, 1982) and for dating works in (Smith and Kelly, 2002). It is calculated as the ratio between the number of unique word types and the total number of tokens in the given text (eq.3). Therefore, a higher lexical density would indicate a wider range of used vocabulary.

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

However, as lexical density counts morphological variants of the same word as different word types, Corpas Pastor et al. (2008) suggested that instead of lexical density, another measure – lexical richness, should be used as an indicative of the vocabulary variety. The lexical richness is computed as the ratio between the number of unique lemmas and the total number of tokens in the given text (eq.4).

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

This second measure does not take into account different morphological counts of the same word as different word types and therefore, Corpas Pastor et al. (2008) believed that it would be a more appropriate indicative of the vocabulary variety of an author.

The second set of features contains nine different part-of-speech frequencies:

- Nouns (N)
- Pronouns (PRON)
- Determiners (DET)
- Prepositions (PREP)
- Adjectives (A)
- Adverbs (ADV)
- Coordinating conjunctions (CC)

[2] http://icame.uib.no/brown/bcm.html

[3] http://khnt.aksis.uib.no/icame/manuals/index.htm

- Subordinating conjunctions (CS)
- Verbs (V)
- Present participles (ING)
- Past participles (EN)

The third set of features were the following 123 stop words (Table 2), based on the 'Default English stopwords list'[4]. In our case, as the used parser treats negative contractions as separate words, transforming for instance *couldn't* into two words: *could* and *not*, we excluded all the words with negative contractions from the original list.

a, about, above, after, again, against, all, am, an, and, any, are, as, at, be, because, been, before, being, below, between, both, but, by, could, did, do, does, doing, down, during, each, few, for, from, further, had, has, have, having, he, her, here, hers, herself, him, himself, his, how, i, if, in, into, is, it, its, itself, me, more, most, my, myself, no, nor, not, of, off, on, once, only, or, other, ought, our, ours, ourselves, out, over, own, same, she, should, so, some, such,than, that, the, their, theirs, them, themselves, then, there, these, they, this, those, through, to, too, under, until, up, very, was, we,were, what, when, where, which, while, who, whom, why, with,would, you, your, yours, yourself, yourselves.

Table 2: Stop words

2. Related Work

Since the 1990s, when the FLOB and Frown corpora were compiled, a great amount of diachronic studies in both American and British English has been conducted using the 'Brown family' of corpora (Brown, Frown, LOB and FLOB). Mair et al. (2003) investigated diachronic shifts in part-of-speech frequencies in British English, reporting an increase in the use of nouns and a decrease in the use of verbs in the Prose text category in the period 1961–1991. Štajner and Mitkov (2011) compared the diachronic changes in the period 1961–1991/2 between British and American language varieties, taking into account four stylistic features: average sentence length (ASL), Automated Readability Index (ARI), lexical density (LD) and lexical richness (LR). Their results indicated increased text complexity (ARI) in the Prose genres of British English, and increased lexical density and lexical richness in the Prose genres of both language varieties over the observed period (1961–1991/2).

It is important to emphasise that in all of these previous diachronic studies conducted on the 'Brown family' of corpora, the authors did not differentiate across different genres in the Prose category (among which is the relevant genre D – Religion), but they rather examined the whole Prose text category together. As the Prose category is comprised of five rather different text genres (Table 1 in Section 1.1), we cannot know whether their findings would stand for the Religious texts (genre D) on its own. Therefore, we included all of these features in our study. This way, by comparing our results with those reported by Štajner and Mitkov (2011), we will also be able to examine whether the religious text had followed the same trends of diachronic stylistic changes as the broader text category they belong to (Prose).

[4]http://www.ranks.nl/resources/stopwords.html

3. Methodology

As some of the corpora were not publicly available in their tagged versions, we decided to use the raw text version of all corpora and parse it with the state-of-the-art Connexor's Machinese syntax parser[5], following the methodology for feature extraction proposed by Štajner and Mitkov (2011). We agree that this approach allows us to have a fairer comparison of the results among different corpora and to achieve a more consistent, highly accurate sentence splitting, tokenisation, lemmatisation and part-of-speech tagging. As the details of tokenisation and lemmatisation process of this parser (Connexor's Machinese syntax parser) were already discussed in detail by Štajner and Mitkov (2011), the focus in this study will be on the POS tagging.

3.1. Part-of-speech tagging

Connexor's Machinese Syntax parser reported the POS accuracy of 99.3% on Standard Written English (benchmark from the Maastricht Treaty) and there was no ambiguity Connexor (2006). For each known word the parser assigns one of the 16 possible morphological (POS) tags: N (noun), ABBR (abbreviation), A (adjective), NUM (number), PRON (pronoun), DET (determiner), ADV (adverb), ING (present participle), EN (past participle), V (verb), INTERJ (interjection), CC (coordinative conjunction), CS (subordinate conjunction), PREP (preposition), NEG-PART (negation particle *not*), INFMARK (infinitive marker *to*).

Here it is important to note that Connexor's Machinese parser differentiate between present and past participle (ING and EN), and verbs (V). This should be taken into account later in Section 4, where diachronic changes of the POS frequencies are presented and discussed. It is also important to emphasise that in the newer version of the parser, the EN and ING forms, which can represent either present and past participle or corresponding nouns and adjectives, are assigned a POS tag (EN, ING, N or A) according to their usage in that particular case. For example, in the sentence:

> *"Some of the problems were reviewed yesterday at a meeting in Paris..."* (LOB:A02),

the word *meeting* was assigned the N tag, while in the sentence:

> *"... Mr. Pearson excels in meeting people informally... "* (LOB:A03),

the same word *meeting* was assigned the ING tag. Similarly, the word *selected* was assigned the A tag in the sentence:

> *"The editors ask some 20 to 30 working scientists to report on the progress made in selected and limited fields... "* (LOB:C14),

[5]www.connexor.eu

while in the other sentence:

> *"... Miss Anna Kerima was selected as best actress... "* (LOB:C02),

the same word *selected* was assigned the EN tag.

3.2. Feature Extraction

All features were separately calculated for each text in order to enable the use of statistical tests of differences in means (Section 3.3) The first four features (ASL, ARI, LD, LR) were computed using the formulas given in Section 1.2. The second set of features (POS frequencies) were calculated separately for each POS tag and for each text, as the total number of that specific POS tag divided by the total number of tokens in the given text (eq.5).

$$< POS > = \frac{total\_number\_of\_ < POS >}{total\_number\_of\_tokens} \quad (5)$$

Stop words were calculated in a similar way. For each stop word and each text, the corresponding feature was calculated as the total number of repetitions of that specific stop word divided by the total number of tokens for the given text (eq.6).

$$< STOP > = \frac{total\_number\_of\_ < STOP >}{total\_number\_of\_tokens} \quad (6)$$

3.3. Experimental Settings

We conducted two sets experiments:

- Diachronic changes of style in British English (1961–1991)
- Diachronic changes of style in American English (1961–1992)

For each experiment we calculated the statistical significance of the mean differences between the two corresponding groups of texts for each of the features.

Statistical significance tests are divided into two main groups: parametric (which assume that the samples are normally distributed) and non-parametric (which does not make any assumptions about the sample distribution). In the cases where both samples follow the normal distribution, it is recommended to use parametric tests as they have greater power than the non-parametric ones. Therefore, we first applied the Shapiro-Wilk's W test (Garson, 2012a) offered by SPSS EXAMINE module in order to examine in which cases/genres the features were normally distributed. This test is a standard test for normality, recommended for small samples. It shows the correlation between the given data and their expected normal distribution scores. If the result of the W test is 1, it means that the distribution of the data is perfectly normal. Significantly lower values of W (≤ 0.05) indicate that the assumption of normality is not met.

Following the discussion in (Garson, 2012b), in both experiments we used the following strategy: if the two data sets we wanted to compare were both normally distributed we used the t-test for the comparison of their means; if at least one of the two data sets was not normally distributed, we used the Kolmogorov-Smirnov Z test for calculating the statistical significance of the differences between their means. Both tests were used in their two independent sample versions and the reported significance was the two-tailed significance. After applying the statistical tests, we only focused on the features which demonstrated statistically significant change (at a 0.05 level of significance) in the observed period. We presented and discussed those changes in Sections 4.1 and 4.2.

After that, we applied several machine learning classification algorithms in Weka[6](Hall et al., 2009; Ian H. Witten, 2005) in order to see which group of the features would be the most relevant for diachronic classification of religious texts. In order to do so, we first applied two well-known classification algorithms: Support Vector Machines (Platt, 1998; Keerthi et al., 2001) and Naive Bayes (John and Langley, 1995) to classify the texts according to the year of publication (1961 or 1991/2), using all features which reported a statistically significant change in that period. The SVM (SMO in Weka) classifier was used with two different settings. The first version used previously normalised features and the second – previously standardised features. Furthermore, we tried the same classification using all possible combinations of the three sets of features: only the first set (1), only the second set (2), only the third set (3), the first and second set together (1+2), the second and third set (2+3), the first and third set (1+3). Then we compared these classification performances with the ones obtained by using all three sets of features together (1+2+3) in order to examine which features are the most relevant for the diachronic classification of this text genre. The results of these experiments are presented and discussed in Section 4.3.

4. Results and Discussion

The results of the investigation of diachronic stylistic changes in religious texts are given separately for British and American English in the following two subsections and compared in the second subsection. The results of the machine learning classification algorithms for both language varieties and the discussion about the most relevant feature set for diachronic classification are given in the third subsection.

4.1. British English

The results of diachronic changes in religious texts written in British English are given in Table 3. The table contains information only about the features which reported a statistically significant change (sign. ≤ 0.05). The columns '1961' and '1991' contain the arithmetic means of the corresponding feature in 1961 and 1991, respectively. The column 'Sign.' contains the p-value of the applied t-test or, alternatively, the p-value of Kolmogorov-Smirnov Z test (denoted with an '*') for the cases in which the feature was not normally distributed in at least one of the two years (according to the results of the previously applied Shapiro-Wilk's W test as discussed in Section 3.3). Column 'Change' contains the relative change calculated as a percentage of the

[6]http://www.cs.waikato.ac.nz/ml/weka/

feature's starting value in 1961. The sign '+' stands for an increase and the sign '−' for a decrease over the observed period. In case the frequency of the feature was '0' in 1961 and different than '0' in 1991, this column contains value 'NA' (e.g. feature 'whom' in Table 3).

Feature	1961	Sign.	Change	1991
ARI	10.332	0.010	+37.72%	14.229
LD	0.304	0.027	+8.47%	0.329
LR	0.268	0.030	+9.40%	0.293
V	14.087	0.020	−14.32%	12.070
PREP	12.277	0.016	+11.69%	13.712
A	6.843	0.009	+24.37%	8.511
ING	0.957	0.042	+31.38%	1.257
an	0.237	0.008	+52.27%	0.361
as	0.554	0.034	+32.68%	0.736
before	0.090	0.002*	−79.98%	0.018
between	0.062	0.046*	+112.07%	0.132
in	1.781	0.046*	+30.57%	2.325
no	0.230	0.027	−46.37%	0.123
these	0.191	0.017*	−45.70%	0.104
under	0.059	0.046*	−77.80%	0.013
whom	0.000	0.006*	NA	0.044
why	0.054	0.046*	−71.25%	0.015

Table 3: British English (1961–1991)

The increase of ARI (Table 3) indicates that religious texts in British English were more complex (in terms of the sentence and word length) and more difficult to understand in 1991 than in 1961. While in 1961, these texts required an US grade level 10 on average for their comprehension, in 1991, they required an US grade level 14. Also, the increase of LD and LR in this period (Table 3) indicates the usage of much wider and more diverse vocabulary in these texts in 1991 than in 1961.
The results also demonstrated changes in the frequency of certain word types during the observed period. Verbs (excluding the past and present participle forms) were used less in 1991 than in 1961, while the prepositions, adjectives and present participles were more frequent in religious texts written in 1991 than in those written in 1961 (Table 3).
The frequency of certain stop words in religious texts had also significantly changed over the observed period (1961–1991). The most striking is the change in the use of the word 'between' which was used more than twice as much in texts written in 1991 than in those written in 1961. Whether this was the consequence of an increased use of some specific expressions and phrases containing this word, remains to be further investigated. Also, it is interesting to note that the word 'whom' was used not even once in the texts from 1961 while it was considerably often used in the texts from 1991.

4.2. American English

The results of the investigation of diachronic stylistic changes in religious texts written in American English are given in Table 4, using the same notation as in the case of British English.
The first striking difference between diachronic changes reported in British English (Table 3) and those reported in

Feature	1961	Sign.	Change	1991
N	27.096	0.009	+10.72%	30.002
PRON	7.763	0.046*	−30.04%	5.431
A	8.092	0.040	+19.63%	9.680
ADV	6.050	0.020	−14.58%	5.168
all	0.298	0.017*	−36.99%	0.188
have	0.500	0.010	−36.52%	0.317
him	0.240	0.017*	−86.25%	0.033
it	0.871	0.015	−32.79%	0.586
not	0.661	0.046*	−13.37%	0.572
there	0.152	0.017*	−42.42%	0.088
until	0.052	0.017*	−60.14%	0.021
what	0.237	0.046*	−9.26%	0.215
which	0.054	0.046*	−48.63%	0.028

Table 4: American English (1961–1992)

American English (Table 4) is that in American English none of the four features of the first set (ASL, ARI, LD, LR) demonstrated a statistically significant change in the observed period (1961–1992), while in British English three of those four features did. Actually, the only feature which reported a significant change in both language varieties during the period 1961–1991/2 is the frequency of adjectives (A), which had increased over the observed period in both cases. This might be interpreted as a possible example of Americanisation – "the influence of north American habits of expression and behaviour on the UK (and other nations)" (Leech, 2004), in the genre of religious texts.
On the basis of the results presented in Table 4, we observed several interesting phenomena of diachronic changes in American English. The results reported a significant increase of noun and adjective frequency, and a significant decrease of pronoun frequency. These findings are not surprising, given that adjectives usually play the function of noun modifiers (Biber et al., 1999) and therefore, an increase of noun frequency is expected to be followed by an increase of adjective frequency. Also, as pronouns and full noun phrases usually compete for the same syntactic positions of subject, object and prepositional complement (Hudson, 1994; Biber et al., 1999), a noun increase and pronoun decrease are not unexpected to be reported together (Mair et al., 2003).

4.3. Feature Analysis

As it was discussed previously in Section 3.3, we used machine learning classification algorithms in order to examine which set of features would be the most important for diachronic classification of religious texts in 20th century. We tried to classify the texts according to the year of publication (1961 and 1991 in the case of British English, and 1961 and 1992 in the case of American English), using all possible combinations of these three sets of features: (1), (2), (3), (1)+(2), (1)+(3), (2)+(3), and compare them with the classification performances when all three sets of features are used (1)+(2)+(3). The main idea is that if a set of features is particularly important for the classification, the performance of the classification algorithms should significantly drop when this set of features is excluded. All experiments were conducted using the 5-fold cross-validation

with 10 repetitions in Weka Experimenter.
The results of these experiments for British English are presented in Table 5. Column 'Set' denotes the sets of features used in the corresponding experiment. Columns 'SMO(n)', 'SMO(s)' and 'NB' stand for the used classification algorithms – Support Vector Machines (normalised), Support Vector Machines (standardised) and Naive Bayes, respectively. The results (classification performances) of each experiment were compared with the experiment in which all features were used (row 'all' in Table 5), using the two-tailed paired t-test at a 0.05 level of significance provided by Weka Experimenter. Cases in which the differences between classifier performance of the particular experiment was significantly lower than in the experiment where all features were used are denoted by an '*'. There were no cases in which a classifier's accuracy in any experiment outperformed the accuracy of the same classifier in the experiment with all features.

Set	SMO(n)	SMO(s)	NB
all	90.19	92.52	85.48
(1)	68.81*	64.05*	67.86*
(2)	68.19*	70.14*	70.33*
(3)	87.24	92.67	88.81
(2)+(3)	91.14	93.33	86.48
(1)+(3)	87.38	89.52	88.43
(1)+(2)	74.48*	66.71*	70.33*

Table 5: Diachronic classification in British English

From the results presented in Table 5 we can conclude that in British English, the changes in the frequencies of the stop words (third set of features) were the most important for this classification task. All experiments which did not use the third set of features (rows '(1)', '(2)' and '(1)+(2)' in Table 5), reported a significantly lower performances of all three classification algorithms.
In the case of American English, we needed to compare only the results of the experiment in which all features were used (row 'all' in Table 6) with those which used only the second (POS frequencies) or the third (stop-words) set of features, as none of the features from the first set (ASL, ARI, LD and LR) had demonstrated a significant change in the observed period 1961–1992 (Table 4, Section 4.2). The results of these experiments are presented in Table 6.

Set	SMO(n)	SMO(s)	NB
all	73.67	77.86	72.90
(2)	70.57	67.67	71.10
(3)	74.67	76.24	78.33

Table 6: Diachronic classification in American English

In diachronic classification of religious texts in American English, no significant difference was reported in the performance of the classification algorithms between the experiment in which both sets of features (POS frequencies and stop-words) were used and those experiments in which only one set of features (either POS frequencies or stop-words) was used. Therefore, based on the results of these experiments (Table 6) we were not able to give a priority to any of these two sets of features in the diachronic classification task.
The comparison of the results of diachronic classification between British and American English (Table 5 and Table 6) lead to the conclusion that the stylistic changes in religious texts were more prominent in British than in American English, as all three classification algorithms in British English (row 'all' in Table 5) outperformed those in American English (row 'all' in Table 6). This conclusion is also in concordance with the comparison between British and American diachronic changes based on the relative changes of investigated features reported in Tables 3 and 4 (Sections 4.1 and 4.2).

5. Conclusions

The presented study offered a systematic and NLP oriented approach to the investigation of style in 20th century religious texts. Stylistic features were divided into three main groups: (ASL, ARI, LD, LR), POS frequencies and stop-words frequencies.
The analysis of diachronic changes in British English in the period 1961–1991 demonstrated significant changes of many of these features over the observed period. The reported increase of ARI indicated that religious texts became more complex (in terms of average sentence and word length) and more difficult to read, requiring a higher level of literacy and education. At the same time, the increase of LD and LR indicated that the vocabulary of these texts became wider and richer over the observed period (1961–1991). The investigation of the POS frequencies also demonstrated significant changes, thus indicating that 30 years time gap is wide enough for some of these changes to be noticed. The results reported a decrease in verb frequencies (excluding the present and past participles) and an increase in the use of present participles, adjectives and prepositions. The analysis of the stop-words frequencies indicated a significant changes in the frequency of ten stop-words (an, as, before, between, in, no, these, under, whom, why) with the most prominent change in the case of the word 'between'.
The results of the machine learning experiments pointed out the third set of features (stop-words frequency) as the most dominant/relevant set of features in the diachronic classification of religious texts in British English. These results were in concordance with the results of the statistical tests of mean differences which reported the highest relative changes exactly in this set of features.
The investigation of stylistic features in 20th century religious texts in American English reported no significant changes in any of the four features of the first set (ASL, ARI, LD, LR) in the observed period 1961–1992. The analysis of the second set of features (POS frequencies) indicated an increase in the use of nouns and adjectives, and a decrease of pronoun and adverb frequencies. In the third set of features (stop-words frequencies), nine words reported a significant decrease in their frequencies (all, have, him, it, not, there, until, what, which). The machine learning experiments, which had the aim of pointing out the most relevant set of stylistic features for diachronic classification of religious texts, did not give the preference to any of the two

sets of features (POS and stop-words frequencies) in this task.

The comparison of diachronic changes in the period 1961–1991/2 between British and American English indicated very different trends of stylistic changes in these two language varieties. From all 146 investigated features, only one feature (adjective frequency) reported a significant change (increase) in both – British and American English, during the observed period (1961–1991/2). The overall relative change was much higher in British than in American English, which was additionally confirmed by the results of the machine learning classification algorithms.

6. References

Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *The Longman Grammar of Spoken and Written English*. London: Longman.

Connexor, 2006. *Machinese Language Analysers. Connexor Manual*.

Gloria Corpas Pastor, Ruslan Mitkov, Afzal Naveed, and Pekar Victor. 2008. Translation universals: Do they exist? A corpus-based NLP study of convergence and simplification. In *Proceedings of the AMTA*.

Nelson W. Francis. 1965. A standard corpus of edited present-day American english. *College English*, 26(4):267–273.

Michael Gamon. 2004. Linguistic correlates of style: authorship classification with deep linguistic analysis features. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

David G. Garson. 2012a. Testing of assumptions: Normality. Statnotes: Topics in Multivariate Analysis.

David G. Garson. 2012b. Tests for two independent samples: Mann-Whitney U, Wald-Wolfowitz runs, Kolmogorov-Smirnov Z, & Moses extreme reactions tests. Statnotes: Topics in Multivariate Analysis.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11:10–18, November.

David Holmes. 1994. Authorship attribution. *Computers and the Humanities*, 28:87–106. 10.1007/BF01830689.

David I. Holmes. 1998. The evolution of stylometry in humanities scholarship.

Richard Hudson. 1994. About 37% of word-tokens are nouns. *Language*, 70(2):331–339.

Eibe Frank Ian H. Witten. 2005. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann Publishers.

George H. John and Pat Langley. 1995. Estimating continuous distributions in Bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345.

S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy. 2001. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13(3):637–649.

Peter J. Kincaid and Leroy J. Delionbach. 1973. Validation of the Automated Readability Index: A follow-up. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 15(1):17–20.

Geoffrey Leech and Nicholas Smith. 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.

Geoffrey Leech and Nicholas Smith. 2009. Change and constancy in linguistic change: How grammatical usage in written English evolved in the period 1931-1991. *Language and Computers*, 69(1):173–200.

Geoffrey Leech. 2004. Recent grammatical change in English: data, description, theory. *Language and Computers*, 49(1):61–81.

Christian Mair, Marianne Hundt, Geoffrey Leech, and Nicholas Smith. 2003. Short term diachronic shifts in part of speech frequencies: A comparison of the tagged LOB and FLOB corpora. *International Journal of Corpus Linguistics*, 7(2):245–264.

Douglas R. McCallum and James L. Peterson. 1982. Computer-based readability indexes. In *Proceedings of the ACM '82 conference*, ACM '82, pages 44–48, New York, NY, USA. ACM.

John C. Platt. 1998. Fast training of Support Vector Machines using Sequential Minimal Optimization. *Advances in Kernel Methods - Support Vector Learning*, 208:1–21.

R. J. Senter and E. A. Smith. 1967. Automated Readability Index. Technical report, Defense Technical Information Center. United States.

J. Smith and C. Kelly. 2002. Stylistic constancy and change across literary corpora: Using measures of lexical richness to date works. *Computers and the Humanities*, 36:411–430. 10.1023/A:1020201615753.

Louis Ule. 1982. Recent progress in computer methods of authorship determination. *Association of Literary and Linguistic Computing Bulletin*, 23(2):73–89.

Sanja Štajner and Ruslan Mitkov. 2011. Diachronic stylistic changes in British and American varieties of 20th century written English language. In *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage at RANLP 2011*, pages 78–85.