ԼԵԶՎԱԲԱՆՈՒԹՅՈՒՆ / LINGUISTICS

Susanna BAGHDASARYAN∗
Astghik MANUKYAN∗∗
Armenian State Pedagogical University after Khachatur Abovyan

THE MOST COMMONLY USED POLITICAL WORDS IN BRITISH AND AMERICAN NEWS REPORTS (CORPUS-DRIVEN ANALYSIS)

It is generally accepted by many scholars that British and American English vary greatly. This is not just a simple assumption about the phenomenon, but rather a well-grounded opinion. On the basis of extensive analysis conducted on differences between the two variations of English, linguists assume that the two vary mainly in grammar patterns and vocabulary choice. This article addresses the question of differences in vocabulary choice focusing on political vocabulary. It compares the most commonly used political words in British and American news reports and reveals distinct contrasts concerning word choice. A comparative analysis with the help of the corpus-driven approach has been chosen as the best procedure to investigate the material under study and demonstrate the mentioned differences.

Key words: political vocabulary, corpus linguistics, corpus-driven approach, authentic text, data sampling, balance, comparative analysis

Introduction

It is common knowledge that British and American English vary greatly, especially in terms of word choice. Fares (2019) argues that lexicon or vocabulary, as one of the central linguistic branches, is an area where apparent variations between British English and American English can definitely be observed. Assumingly, the political vocabulary of the two varieties will differ as well, which will be found out by "posing" corpus

---

∗ susanna.baghdasaryan@gmail.com
∗∗ astghikmanukyan1996@gmail.com

linguistics a question to answer. Corpus linguistics, being a relatively new branch of linguistics, deals with the language as it is without prescribing any rules to it. Until very recently, linguists mainly referred to corpus linguistics as a separate discipline of linguistics, as an objective per se (McCarthy & O'Keeffe, 2010) while now the tendency is to regard it as a tool with the help of which many academic disciplines can be developed. It should be noted here that it can be applied in literary and translation studies, language teaching and learning, discourse analysis, and the list still can be continued. Corpus linguistics can be beneficially applied in different areas of language analysis, including the vocabulary of a certain field such as politics.

Nowadays, an average person has unlimited access to political news with the help of the Internet and online news webpages (Kahin & Varian, 2000). Many world-renowned news sites such as BBC, CNN, Daily Mail, The Times enable millions of people from all around the globe to be informed on the spot. It is undeniable that "the contemporary world has realized that the media, particularly newspapers, have an indispensable role in political life" (Kahin & Varian, 2000).

When it comes to the news coverage, the choice of political vocabulary is of high importance. The words the newspaper writers or editors use have a huge impact on the audience and may somehow "shape" the opinion of the readers (Kutter & Kantner, 2012). From country to country, the choice of vocabulary may vary (Edelman, 1977). We may assume that even two English speaking communities - the UK and the USA - use different vocabulary when writing political news reports. The factors listed below may have an impact on the vocabulary choice:

- the geographical dimensions of a particular country,
- the time of publication,
- the political situation of a particular country,
- the government type (monarchy, federal government),
- the relationship with other countries.

Corpus-driven Method and its Contribution to Comparative Analysis of the Most Commonly Used Political Words in British and American News Reports

When we speak about corpus linguistics, it is highly important to mention that there are two approaches - corpus-based and corpus-driven, and these two are considered to be somehow opposed to each other. Tognini-Bonelli (2001, p. 65) discusses the difference between these two approaches in her book

Corpus Linguistics at Work. According to her, "the term 'corpus-based' is used to refer to methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study". To put it simply, we can say that corpus-based studies, as opposed to corpus-driven studies, mainly utilize data extracted from corpus for the purposes of exploring a theory or hypothesis, aiming to validate, disprove or refine it. Meanwhile, the corpus-driven approach is regarded as a methodology where the corpus serves as an empirical basis, as an organized collection of representative data from which linguists extract data and detect linguistic phenomena without having any prior assumptions and expectations. As a matter of fact, linguists solely pose a question to the corpus and follow what the answer will be (Tognini-Bonelli, 2001). Any inferences or presumptions are made mainly on the basis of data extracted from the source corpus (Storjohann, 2005). According to the corpus-driven approach, the characterization of corpus linguistics as a method is not accepted. On the contrary, corpus is regarded as an exhaustive source of information that is highly representable data (data-driven approach), hence it can serve as a basis for our hypotheses about language. Having mentioned the differences between the two approaches, we can sum up by quoting a sentence by Tognini-Bonelli (2001, pp. 84-85) who recapped in this way; "the corpus itself embodies a theory of language".

For the purposes of answering the above-mentioned research question, a corpus-driven method of analysis has been used. A specialized, small corpus has been compiled to find out the most common political words in the British and American news reports.

As Nelson (2010) states, to have a real corpus, there are certain steps and procedures to follow. Here we should be very careful as the corpus is not simply a collection of the texts (Biber & Reppen, 1998). So, in order to have a real corpus and not a mere compilation of texts, an explicit and distinct strategy needs to be created long before anything else can be done.

Koester (2010) is one who mentioned the importance of sampling, balance and representativeness, which are tightly interrelated. If we are to describe them, we should say that sampling is deciding the type and genre of the text that is going to be included in the corpus (in this case news reports). Sampling is of great importance and serves as a ground for the other two, and, as a ground, requires a lot of attention, because there are no certain lines that differentiate one genre or text type from the other. Besides, the compiler should also pay attention to representativeness and balance of samples included (e.g. the length of the text, text type) so to avoid biased results.

5

In our case, it has been decided to take the samples of the corpus from ten different online pages - five for the UK and five for the USA respectively. From each source webpage, fifteen sample articles have been chosen with an average size of five hundred words. To find the most famous British news webpages, Feedspot blog has been used and according to it, BBC is on the top of the UK list followed by Daily Mail Online, The Telegraph, The Independent, Daily Mirror, The Sun and many others. The news reports were taken from BBC, The Daily Mail, The Independent, Daily Mirror and The Telegraph and a small corpus for UK news webpages in size of 61.816 words has been built. The Feedspot blog provided the list of famous US news sites as well. The forefront of this list is CNN, after which The New York Times, USA Today, Los Angeles Times, Chicago Tribune, Newsmax and The Washington Post are placed as sources of sample texts. Similarly, seventy-five news reports have been taken from these widely-spread sources, fifteen texts for each of them. For US news reports we have seventy-five files containing 61.269 words. Overall, the two taken together include 150 sample texts with 123.085 words.

The availability of computerized text samples is likely to be a successful source for text-analytical research. Yet, a lot more attention should be paid to the cleaning of so-called "noises" that can induce biases and distort the reliability of our analysis. As Gabrielatos (2007, p. 6) states "even a properly balanced corpus of texts, e.g. balanced with the regard to time-span, sources, languages, or wording related to the issue under investigation, may be incomplete and contain a lot of 'noise'". Hence, a careful consideration during data collection needs to be done if we want to avoid biased results.

After the choice of relevant sources and sample texts, they have been copied and pasted in word documents. The Antconc tool which was used for this research can only analyze the texts as long as they are in UTF 8 format. So, the collected texts were saved as ordinary texts in UTF8 format. In order to keep away from biased results, all the unnecessary data- noises were carefully removed from the documents. As soon as the text files were ready, the second step could be processed and a research process with the help of the Antonc Tool might be commenced.

Basically, "Antconc is a freeware, multiplatform tool for carrying out corpus linguistics research and data-driven learning" (Anthony, 2014, p. 7). This software contains several tools such as Concordance Tool, Clusters/N-Grams, Collocates, Word List as well as Keyword List. For this particular research the Wordlist Tool was used as it "counts all the words in the corpus and presents them in an ordered list" and "this allows to quickly find which words are the most frequent in a corpus" (Anthony, 2014, p. 10). As we are

to find the most frequently used words, the Word List Tool is a perfect match. By means of this tool all the words in the corpus have been counted and presented in an ordered list. This enables us to quickly find the words which are the most frequent in a corpus. For the vividness and clearness of results, the investigations of British and American news reports have been presented separately.

The results from the UK news reports: To begin with, we should state that when analyzing the results gained during this research it is observed that there is a certain group of words that constitutes a stable political vocabulary and there are words that come along in a course of certain circumstances that affect the political situation of the country at that moment. The most commonly used political vocabulary found in British news reports is presented below. The screenshot will vividly depict the results. The list starts with the word party (political party) which is found 260 times. Right after this, as expected the word Brexit (223 times) comes, which is, in fact the withdrawal of the UK (occurring in corpus 203 times) from the EU, i.e. European Union (103 times) and which was then a topic of heated discussions and debates in UK.

Picture 1

When it comes to the popularity of the UK political parties, the Labour party appears in the forefront (153 times), as compared with Tory (73 times), Conservative party (50 times), UKIP – UK Independence Party (51 times)

and Liberal party (33 times). As the corpus is a rather small one, everything that appeared more than 45 times is considered to be commonly used. See the results in the table below.

Table 1

The most frequently occurring words	
Word	**Frequency**
Government	167
Deal	100
Minister	100
Leader	84
Public	81
Political	77
President	75
Secretary	56
Tax	57
Referendum	49
Coalition	36
Candidate	35
Policy	43
Debate	27

Now, let us turn our attention to the recent events that highly affected our results. An election to the European Parliament was held between 23 and 26 May 2019, which is, in fact, the ninth parliamentary election. Fifty sample texts out of seventy-five were gathered during the period of parliamentary elections that is why we have such an "active" election-related vocabulary.

Table 2

Elections-related vocabulary	
Word	**Frequency**
MP (member of parliament)	76
Vote	70
election/elections	127
Parliament	58
Campaign	32
seat/seats	46

In the above mentioned table (Table 2) the word seats was taken as a political word as this word represents "a position as an elected member of a government, or as a member of a group that makes official decisions". In order to see the usage of that word in the corpus, a screenshot is provided.

Picture 2

The results from the US news reports: The results gained from the US news+ reports corpus are interesting as well. The most common political word there is the word president occurring 248 times.

Picture 3

When we consider the political parties of the US it is important to mention that the Democrats (81) are winning in terms of popularity and amount of news coverage, then comes Republicans (80 times). The table below shows the most common political words for the US news reports. The words mentioned above are already omitted from the table.

9

Table 4

The most frequently occurring words	
Word	**Frequency**
tariffs	131
campaign	87
democratic	87
federal	77
government	70
senate	69
policy	57
committee	55
congress	49
vote	59
party	47
war	44
election	40
impeachment	39
presidential	39
senator	26

Currently, the US is suffering hard because of the number of immigrants from Mexico who are trying hard to cross the border of the "dreamland" United States of America. The US government is implementing all the possible forces to escape from having Mexican immigrants, and it is becoming too hard to deal with them. Travel administration under the President's guidance has decided to limit the unnecessary travel with Mexico. That is why we see the name of Mexico 133 times in the corpus. The other words concerning the issue of immigration are border (89), Mexican (75), migrants (50), immigration (37).

Table 5

Migration issues vocabulary	
Word	**Frequency**
border	89
Mexican	75
migrants	50
immigration	37

Nowadays, the learner-centered approach to teaching is becoming excessively popular, and as Brown (2003, p. 49) has stated; "The premise 'one teaching style fits all', which is attributed to a teacher-centered instructional approach, is not working for a growing number of diverse, student populations." From the perspective of the latter-mentioned approach, the teaching-learning process is regarded as a process that is designed to meet the needs of the learners, so the teaching materials should also address those needs. Course materials should help the learners to overcome the difficulties that they will probably face in a real classroom-cut environment, especially when interacting with native speakers. Rogers and Medley (1988), in their turn, emphasized the importance of giving the learner the chance of developing and assimilating the needed strategies which will enable them to deal with the actual language used outside the comfortable premises of the classroom. And as Nadrag (2017, p. 2) says the learners are particularly "fascinated by the use of authentic materials, especially if these materials refer to topics they are interested in" as those materials "offer them the opportunity to further inform and use English in their day-to-day activities and enable them to communicate their thoughts with confidence whether inside or outside the classroom." This is how the issue of authenticity comes forward, as authentic materials are said to "link the formal, and to some extent artificial environment of the classroom with the real world" in which the learners "will eventually be using the language they are learning" (House, 2008). A rather comprehensive definition of authenticity is provided by Bacon and Finnemann (1990, pp. 459-473) who state that "authentic materials are those texts which are made by native speakers for non-pedagogical purposes".

Having defined what authenticity is and having stressed out the importance of using authentic material, let me state here that the results of the following research can be beneficially used in the teaching process, especially while teaching political vocabulary. The findings represent the actual political vocabulary of the two varieties of English and as the text samples taken for the corpus compilation process are authentic (texts were published originally in news webpages by native speakers and they carry only informative function), they simply depict the actual language, or to put it in other words "the language as it is".

Corpus linguistics is a relatively new branch of linguistics that descriptively deals with the actual language in order to examine the language as it is and to find the rules that govern the language. In fact, opinions about corpus linguistics differ; one group of linguists regard it as an end itself, while others see the corpus as a means to an end and thus, the

linguists of the second group apply it in areas such as language teaching and learning, discourse analysis, literary stylistics, forensic linguistics, pragmatics, speech technology and sociolinguistics. The opposition of these two opinions served as a basis for two opposite approaches: the corpus-based and the corpus-driven. The former can be defined as a method using corpus as an inventory of language data (Storjohann, 2005), while corpus-driven linguistics rejects the characterization of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language (Tognini-Bonelli, 2001). For this article, the corpus-driven approach has been considered due to the need to create a small specialized written corpus.

Conclusion

The news reports' political vocabulary of two English-speaking communities - British and American - has been investigated and compared and the most frequently used political vocabulary in the UK and the US has been singled out to see whether or not they contain similar words and to put the gained knowledge into usage. Two small specialized corpora - one for the British and the other for the American news reports have been compiled. The results show that although the people of these countries speak the same language, they speak differently and the reasons depend on the politics of the county, geographical dimension, political situation, economic situation, government type, time of publications and many others. The results from both corpora show the following:

The US newspapers write about or at least mention the name of the president more than their British colleagues whose top-used word is (political) party.

By the time of corpus compilation, the UK was busy with the EU Parliament elections and Brexit process, while Americans were handling the immigration situation. This is why words like vote, election, parliament, and EU are used more frequently in the UK news reports and words like immigration, border, and migrants are used more frequently in the US news reports. The UK news webpages refer to political parties more than the Americans ones.

Both corpora contain the following political terms as frequently used elements; government, vote, president, right, policy, country, elections.

Finally, it should be noted that the findings retrieved from the authentic texts used in this study, can be used in the teaching-learning process, as they provide the most commonly used political vocabulary in English.

REFERENCES

Anthony, L. (2014, September). AntPConc: A Freeware Multi-Platform Parallel Concordancer. Paper presented at the American Association for Corpus Linguistics (AACL 2014), September 25-28, 2014, Flagstaff, Arizona.

Bacon, S. M., & Finneman, M. D. (1990). *A study of the attitudes, motives, and strategies of university foreign language students and their disposition to authentic oral and written input.* The Modern Language.

Biber, D., Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. *Learner English on computer.*

Brown, K. L. (2003). From Teacher-Centered to Learner-Centered Curriculum: Improving Learning in Diverse Classrooms. *Education, Vol. 124, Issue 1,* 49-54.

Edelman, M. J. (1977). *Political Language: Words That Succeed and Policies That Fail.* University of Wisconsin, Academic Press.

Fares, A. (2019). British English and American English: History and Differences. *International Journal of Linguistics, Literature and Translation (IJLLT).*

Gabrielatos, C. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal 31*, 5-4.

House, S. (2008). Authentic materials in the classroom. *Didactic approaches for teachers of English in an international context,* 53-70.

Kahin, B., & Varian, H. (2000). Profiting from Online News. *Internet Publishing and Beyond; The Economics of Digital Information and Intellectual Property.* MIT Press.

Koester, A. (2010). Building small specialized corpora. *The Routledge of Corpus Linguistics.* Routledge.

Kutter, A. & Kantner, C. (2012). Corpus-based content analysis: a method for investigating news coverage on war and intervention. *International Relations Online Working Paper Series, 2012(01).*

McCarthy, M. & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? *The Routledge of Corpus Linguistics.* Routledge.

Nadrag, L. (2017). *The benefits of using authentic materials in ESP classroom.* Ovidus university press.

Nelson, M. (2010). Building a written corpus: What are the basics? *The Routledge of Corpus Linguistics.* Routledge.

Rogers, C., & Medley, F. (1988). Language with a purpose: Using authentic materials in the foreign language classroom. *Foreign Language Annals, 21,* 467-478.

Storjohann, P. (2005). *Corpus-driven vs. corpus-based approach to the study of relational patterns.* Institut für Deutsche Sprache Mannheim (IDS).

Tognini-Bonelli, E. (2001). Corpus Linguistics at Work. *Studies in Corpus Linguistics, 6.* John Benjamins.

Top 50 UK News Websites. Retrieved from https://blog.feedspot.com/uk_news_websites/

Top 50 USA News Websites. Retrieved from https://blog.feedspot.com /uk_news_websites/

Longman Dictionary of Contemporary English Online. https://www.ldoceonline.com/

SOURCES OF DATA

BBC https://www.bbc.com/news/politics
CNN https://edition.cnn.com/politics
Daily Mirror https://www.mirror.co.uk/news/politics/
Los Angeles Times https://www.latimes.com/
The Daily Mail https://www.dailymail.co.uk/news/uk-politics/index.html
The Independent https://www.independent.co.uk/news/uk/politics
The New York Times https://www.nytimes.com/video/us-politics
The Telegraph https://www.telegraph.co.uk/politics/
The Washington Post https://www.washingtonpost.com/politics/
USA Today https://www.usatoday.com/news/politics/

Ս. ԲԱՂԴԱՍԱՐՑԱՆ, Ա. ՄԱՆՈՒԿՅԱՆ – *Բրիտանական և ամերիկյան մամուլում գեստեղված քաղաքական բառապաշարի համեմատական վերլուծություն կորպուսի հիման վրա.* – Գիտնականների կարծիքով՝ անգլերենի երկու տարբերակները՝ բրիտանական անգլերենը և ամերիկյան անգլերենը, մեծապես տարբերվում են: Լայնածավալ հետազոտություն-ների, վերլուծությունների հիման վրա լեզվաբանները եկել են այն եզրահանգման, որ անգլերենի այս երկու տարբերակները հիմնականում տարբերվում են քերականական օրինաչափությունների և բառապաշարի ընտրության հարցերում: Այս հոդվածը անդրադառնում է բառապաշարի ընտրության և մասնավորապես, քաղաքական բառապաշարի տարբե-րությունններին վերաբերող հարցին: Հոդվածի ուսումնասիրության առարկան բրիտանական և ամերիկյան լրատվական նյութերում

գետեղված առավել հաճախ օգտագործվող քաղաքական բառապաշարն է: Տվյալ ուսումնասիրությունը կատարվել է կորպուսի հիման վրա համեմատական վերլուծության մեթոդի կիրառմամբ:

Բանալի բառեր. քաղաքական բառապաշար, կորպուս լեզվաբանություն, կորպուսի վրա հիմնված ուսումնասիրություն, ընագիր, տվյալների ընտրանք, հավասարակշռություն, համեմատական վերլուծություն

С. БАГДАСАРЯН, А. МАНУКЯН – *Наиболее часто используемая политическая лексика в британских и американских новостях (на основе корпусных данных).* – Многие ученые считают, что два варианта английского языка – британский английский и американский английский – существенно различаются. Следует отметить, что это не голословное предположение, а скорее обоснованное мнение. Основываясь на данные обширного анализа, проведенного при исследовании различий между двумя вариантами английского языка, лингвисты пришли к выводу, что они различаются в основном выбором лексики и грамматических моделей. В данной статье рассматривается вопрос о различиях в выборе лексики вообще и политической лексики, в частности. Исследуя наиболее часто используемую политическую лексику в британских и американских новостных сообщениях, авторы попытались выявить определенные различия в выборе лексики, как общей, так и политической. Для проведения исследования был выбран метод сравнительного анализа на основе корпусных данных.

Ключевые слова: политическая лексика, корпусная лингвистика, корпусный подход, аутентичный текст, выбор данных, сбалансированность, сравнительный анализ