IRSTI 11.07.13 https://doi.org/10.26577/jpcp.2020.v73.i3.08

I.A. Bykov

St Petersburg State University, Russia, St Petersburg,
e-mail: i.bykov@spbu.ru

STUDYING POLITICAL DISCOURSE OF THE PRESIDENT ADDRESS IN RUSSIA WITH THE TEXT MINING TECHNIQUE

The article describes the technique and results of the study of political discourse using text mining technology with the statistical package R. Unlike traditional content analysis, text mining uses automated methods for processing text in natural languages. The article presents a specific technique of computational operations and an algorithm of visualization. The study aims to study the corpus of texts of the President Address to the Federal Assembly in Russia. The study describes the evolution of the political agenda in post-Soviet Russia within the discourse analysis approach. The study shows that the idea of 'democracy and human rights' fails to be the key concepts of public policy in Russia. The presidents of Russia usually stick around 3 common topics: Russia, state, and power. The author approves that text mining allows us to automate part of the research work and it functions as one of the directions for a comprehensive analysis of political discourse. The applied technique can be used to automate research in political linguistics, as well as to study different types of political documents and texts.

Key words: text mining, political discourse, president address, political communication, mathematical linguistics.

И.А. Быков

Санкт-Петербург мемлекеттік университеті, Ресей, Санкт-Петербург қ.,
e-mail: i.bykov@sbpu.ru

Ресейдегі президенттік жолдаулардың саяси дискурсын зерттеу мәтінді өндіру әдісін қолдану

Мақалада статистикалық пакеттегі мәтінді іздеу әдісін қолдана отырып, саяси дискурсты зерттеу әдістемесі мен нәтижелерін сипаттауға арналған. Классикалық мазмұнды талдаудан айырмашылығы, мәтінді өңдеу табиғи тілде жасалған мәтінді өңдеудің автоматтандырылған әдістерін қолданады. Мақалада компьютерлік операциялардың арнайы әдістерін тұжырымдау және нәтижелерді визуализациялау берілген. Әдістеме саяси дискурсты зерттеудің заманауи тәсілдерін қолдана отырып, Ресей Федерациясының Федералды жиналысына президенттік жолдаудың мәтіндерін зерделеуге бағытталған. Зерттеу посткеңестік кеңістіктегі Ресейдегі президенттік жолдаудың мысалын қолдана отырып, саяси күн тәртібінің эволюциясын сипаттайды.

Бұл зерттеу демократия және адам құқықтары идеялары Ресейдегі мемлекеттік саясаттың негізгі тұжырымдамаларына айналмағанын көрсетеді. Ресей Федерациясының президенттері әдетте өз сөздерін негізгі үш проблеманың төңірегінде шоғырландырады: Ресей, билік және мемлекет. Президенттік хаттардағы саяси дискурс посткеңестік кеңестік Ресейдегі саясаттың жаппай қабылдауын көрсететіні анық. Біздің зерттеуіміз мәтіндік іздеу зерттеу бағдарламасының бір бөлігін автоматтандыруы мүмкін және саяси дискурсты терең талдау әдісі ретінде қолданыла алатындығын растайды. Қолданбалы әдістемені саяси лингвистикадағы зерттеулерді автоматтандыру үшін, сондай-ақ әр түрлі типтегі саяси құжаттар мен мәтіндерді зерттеу үшін қолдануға болады.

Түйін сөздер: саяси дискурс, президенттік үндеу, саяси коммуникация, математикалық лингвистика, мәтін шығару.

И.А. Быков
Санкт-Петербургский государственный университет, Россия, г. Санкт-Петербург,
e-mail: i.bykov@sbpu.ru

Исследование политического дискурса президентских посланий в России с помощью методики text mining

Статья посвящена описанию техники и результатов исследования политического дискурса с использованием метода text mining в статистическом пакете R. В отличие от классического контент-анализа, text mining использует автоматизированные методы обработки текста, созданные в естественных языках. Статья содержит описание специальной методики компьютерных операций и визуализации результатов. Методика нацелена на изучение текстов президентских посланий Федеральному Собранию России с использованием современных подходов изучения политического дискурса. Исследование описывает эволюцию политической повестки в постсоветской России на примере президентских посланий. Данное исследование показывает, что идеи демократии и прав человека не стали ключевыми концептами публичной политики в России. Президенты Российской Федерации обычно концентрируют свои выступления вокруг трех основных проблем: России, власти и государства. Очевидно, что политический дискурс президентских посланий отражает массовые представления о политике в постсоветской России. Автор утверждает, что text mining позволяет автоматизировать часть исследовательской программы и может быть использован в качестве метода углубленного анализа политического дискурса. Примененная методика может быть также использована для автоматизации исследований в политической лингвистике и изучения разных видов политических документов и текстов.

Ключевые слова: политический дискурс, президентское послание, политическая коммуникация, математическая лингвистика, text mining.

Introduction

The political science as a field of research shares its research methods and techniques with the other social sciences such as sociology, psychology, economics, etc. Content analysis has been an important part of political studies since the early stages of the research. However, today it is a particularly important to know how to use the content analysis for the two specific reasons: (1) there are too many scientific software available all around the world, (2) this method sometimes is the only method possible to use in order to get some reliable data from the closed or non-transparent political systems. This article aims to describe the evolution of political agenda by means of studying the content of President Address in Russia. The study calls to contribute into development of computational linguistic studies of political communication today.

Literature review

In the times of computational studies it is possible to automate many parts of scientific research. In social sciences there is such a well-known piece of software as the SPSS (Statistical Packages for the Social Sciences). Currently, the SPSS belongs to the IBM company and costs over 1000 US dollars per user for 1 year. However, there is an open source alternative for the SPSS available. It is R (https://www.r-project.org/). R is a programming language and a software environment for statistical computing (Kabacoff, 2011). First big advantage of R is about pricing. Being open source project R is absolutely free for all users. Second, there are many special packages for different research techniques. So, R consists of the core programming environment which is good for basic statistical operations and the additional programs for special research techniques such as advanced graphics, network analysis, machine learning, etc. For example, the text mining as a method of automated content analysis with big data is well developed in R (Feldman, Sanger, 2006; Hotho, Nürnberger, Paaß, 2005; Practical Text Mining, 2012). On the other hand, it is worth to mention that there are not so many publications about discourse analysis with R available in Russia. We can only mention the article by A. Nosov (Nosov, 2018). In this particular study the ‘tm-package’ has been used (Feinerer, Hornik, Meye, 2008).

Political science applies discourse analysis in order to study various field of political communication. For example, Balahonskay with colleagues have studied verbal aggression in mass media of Russia with formal linguistic approach (Balahonskaja, Bykov, 2018; Bykov, Balakhonskaya, Gladchenko, Balakhonsky, 2018). The methodology of discourse analysis seems to be sufficiently developed. Among modern foreign scholars studying the problems of discourse theory, one can single out G. Brown, G. Yule, T. van Dijk, N. Fairclough, D. Schiffrin, M. Stubbs, R. Wodak

(Brown, Yule, 1983; Dijk, 2008; Fairclough, 2007; Schiffrin, 1994; M. Stubbs, 1983; Wodak, 1989). In recent years, discourse analysis has significantly expanded its methodological and instrumental base, and its basic principles are successfully applied to the study of various communication practices.

However, we should get into account the fact that there are many techniques of discourse analysis. Maybe, one of the most popular approaches is a critical discourse analysis (Upravljaemost' i diskurs, 2019). Linguistic analysis and other forms of not-automated analysis take the biggest share of discourse studies today. However, to us the non-automated studies produce limited results in terms of objectivity and visualization.

Text mining usually aims to analyze large texts with high level of objectivity. Being quantitative method it uses algorithms to study texts of natural languages as well as of meta-data from all sorts of big data available today from all kinds of data bases and social data-sets. Developing of automated discourse analysis will definitely help to produce better results in political studies.

Method

In this study we run text mining algorithm in 4 versions of the President Address to the Federal Assembly of Russia. These speeches were presented by Boris Yeltsin in 1994, by Vladimir Putin in 2000 and in 2018, and by Dmitry Medvedev in 2008. All texts are available via the official web-site of the President of Russia (www.kremlin.ru) in English. Unfortunately, there is no official translation into English of Yeltin's Address in 1994. In this case we have to work with the version in Russian.

In order to follow the algorithm one should install 3 additional packages in R: 'NLP' (natural language processing), 'tm' (text mining), and 'wordcloud'. Basically, the algorithm consists of 3 stages. During the first stage, one should convert text of the given President Address into the corpus. In the second stage, one should steam the corpus automatically removing non-important symbols and stop-words from the text. The third stage deals with the building of a word-cloud for the most frequently used words.

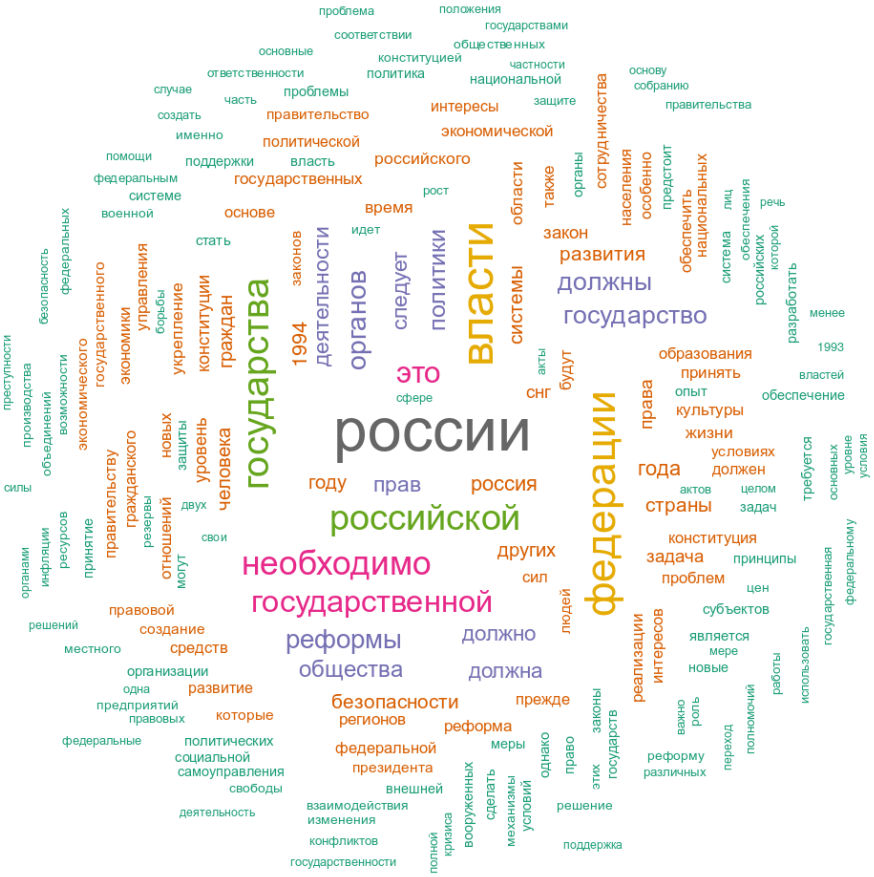

Figure 1 – Word cloud for President Address in 1994 by B. Yeltsin
Notes: composed by the author

Let me illustrate all stages with the coding samples from R. In order to start the algorithm one need to start R and load all necessary packages with this commands:

```
library("NLP")
library("tm")
library("SnowballC")
library("RColorBrewer")
library("wordcloud")
```

All texts should be in plain text format (.txt) which allows us to load text into corpus with this commands:

```
putin <- Corpus(DirSource("~/Putin2000/"))
myCorpus <- Corpus(VectorSource(putin))
```

In this particular example we took text from the folder "Putin2000" and convert a txt-file into the special corpus-file with all characters ordered in vector.

The second stage is about manipulations with corpus. In the beginning, one should remove special characters and replace them with spaces:

```
toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))
myCorpus <- tm_map(myCorpus, toSpace, "/")
myCorpus <- tm_map(myCorpus, toSpace, "@")
myCorpus <- tm_map(myCorpus, toSpace, "\\|")
```

Next step is about the cleaning the text with setting all letters in lower case, removing common stop words in English, removing symbols of punctuation, ets. Here are the commands to do all manipulations

```
myCorpus <- tm_map(myCorpus, tolower)
myCorpus <- tm_map(myCorpus, removeWords, stopwords("english"))
myCorpus <- tm_map(myCorpus, removePunctuation)
myCorpus <- tm_map(myCorpus, stripWhitespace)
myCorpus <- tm_map(myCorpus, stemDocument)
```

And the third stage is about visualization. One need to build a matrix of words with the functions "TermDocumentMatrix" like that:

```
myDtm <- TermDocumentMatrix(myCorpus, control = list(minWordLength = 1))
```

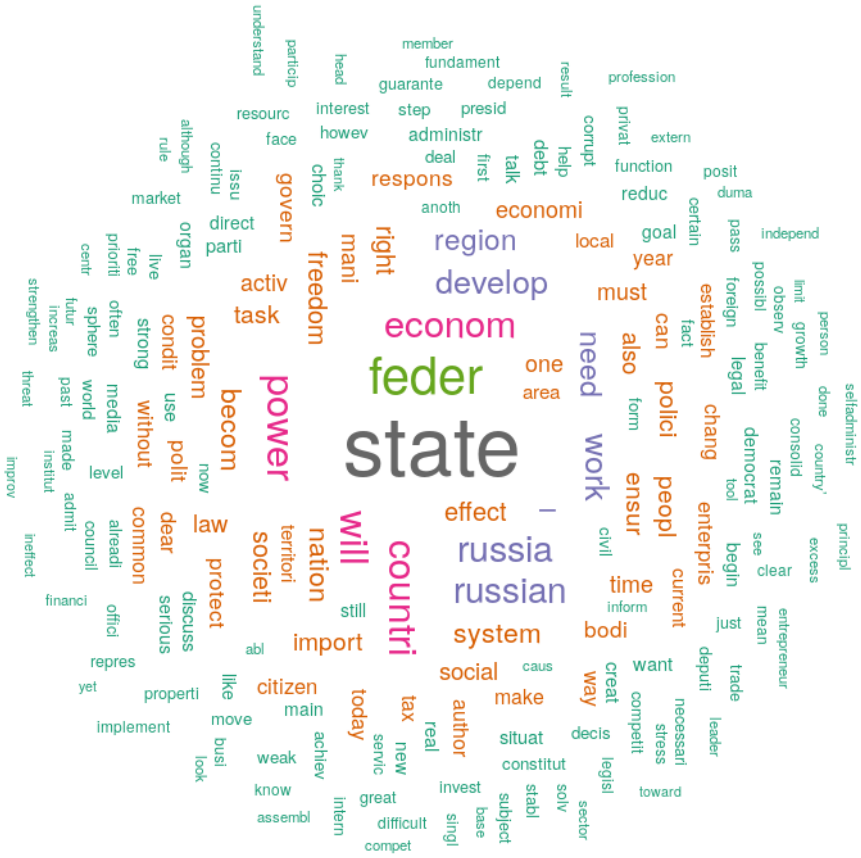

Figure 2 – Word cloud for President Address in 2000 by V. Putin
Notes: composed by the author

```
m <- as.matrix(myDtm)
v <- sort(rowSums(m),decreasing=TRUE)
d <- data.frame(word = names(v),freq=v)
head(d, 10)
```

At last, on should generate a word cloud:

```
set.seed(1234)
wordcloud(words = d$word, freq = d$freq, min.freq = 1,
max.words=200, random.order=FALSE, rot.per=0.35,
colors=brewer.pal(8, "Dark2"))
```

Since visualization is a crucial part of the modern scientific research it is highly important to be able to build a word cloud. Of course, there are many online services for word-cloud, but there is no certainty about the algorithm behind them. All texts, images and R-codings are available in author's github-page (https://github.com/bykov404/txt-mining-rus-ling-bulletin).

Results

President address is a key public relation text in political life of any presidential political system. D. Gavra suggests that a president address is a complicated social phenomenon being a special genre of PR-texts and at the same time it belongs to a much wider communicative continuum rather than exclusively to the sphere of political communication. (Gavra, 2016). As a sociological phenomenon, the presidential address in its most general form can be defined as a complex, controversial social phenomenon that has dual nature and special functions in society. Dual nature deals with the ambiguity of possible ways of social being of the address. On the one hand, president address has spiritual essence. On the other hand, it usually has practical purpose.

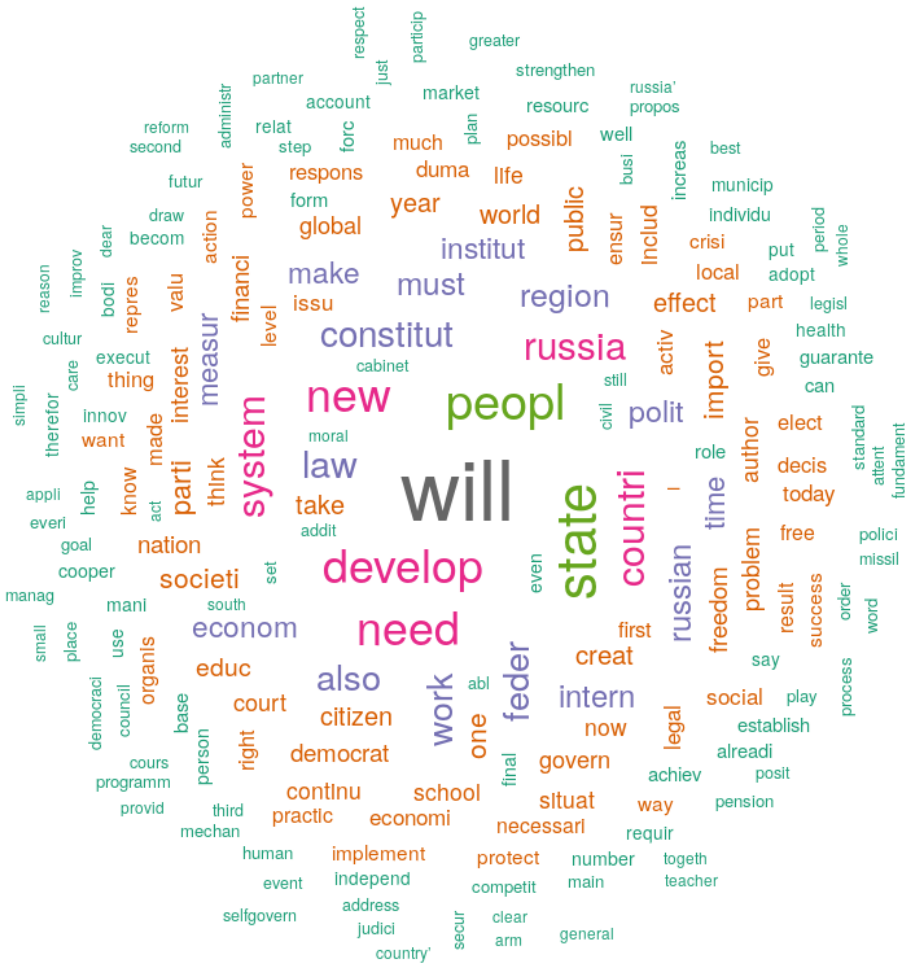

Figure 3 – Word cloud for President Address in 2008 by D. Medvedev
Notes: composed by the author

So, it is not a big surprise that in Russia the president addresses by Yeltsin, by Medvedev, and especially by Putin have been studied well enough in previous studies (Vasilev, 2015). Gavrilova claims that linguistic analysis of the presidential speeches has a good tradition in political linguistics (Gavrilova, 2016). V. Kosov applied comparative approach to study President Addresses in Russia and France (Kosov, 2019). Patocka-Sigłowy points out that in his President Address of 2007 'Putin has managed to respond to all the needs of Russian people, who are in need of a "strong hand" in the sense of security and stabilization' (Patocka-Sigłowy, 2015: 27). Unexpectedly, the text mining has not been applied to this issue yet.

Visual results of the text mining are presented in 4 figures below. As you can see, all 4 versions stick around 3 common topics: Russia, state, and power. These topics reflect general political attitude toward stability and order in contemporary public opinion in Russia. However, the other key topics tend to change. Next paragraphs deals with the major topics of the each president address.

In 1994 B. Yeltisn devoted significant part of the speech to problem of state, society, economical reforms, and federalization (see figure 1). Surprisingly, the topic of constitution took very insignificant place in his speech despite the fact that the Constitution of Russia was adopted one year before. Central themes of the Yeltsin's speech were about Russia, power, government, and federation. To Yeltsin federalism was the only way of development for Russia.

In 2000 V. Putin also spoke about state, federalization, and economics (see figure 2). However, the topic of reforms was substituted by the topic of economic development. As Yeltsin, Putin also had a big deal of discussion about federalism. However, unlike Yeltsin Putin wonted to reinforce a vertical of power.

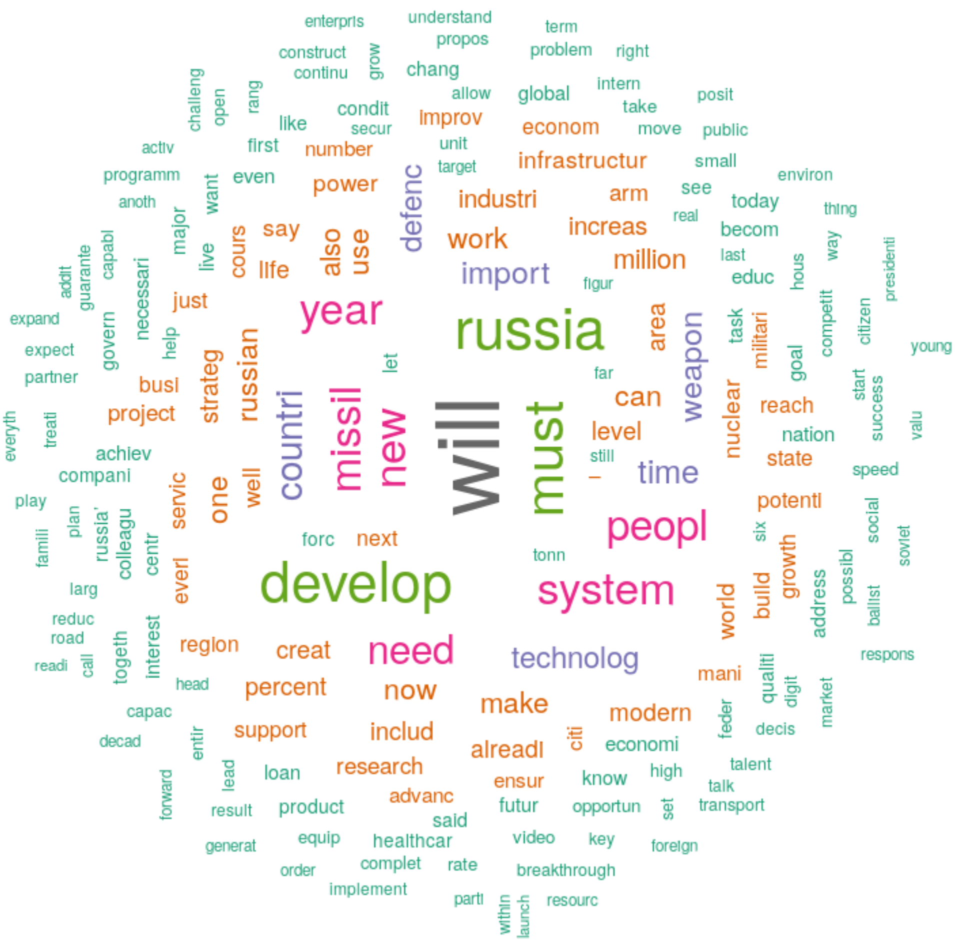

Figure 4 – Word cloud for President Address in 2018 by V. Putin
Notes: composed by the author

In 2008 D. Mevedev included in his speech topics of law and constitution (see figure 3). During his tenure D. Medvedev dedicated a lot of his efforts to the legal system of Russia.

In 2018 V. Putin introduced new topics about weapons, defense, and missiles (see figure 4). This speech reflects tendency toward warfare rhetoric in political communication of Russia in last decade. For the first time in history Putin used a video-clips with 3D-graphics about new weapons of mass destruction which are now available to defend Russia.

Conclusion

As M. Gavrilova has concluded in her summary article, 'it is important to take into consideration not only the semantic topics which *were included* in texts but also the semantic topics which *were not included*' (Gavrilova, 2013: 111). Following this suggestion, we have found that in all 4 texts the topics of democracy and human rights are not visible playing no role in political discourse of the President Address. This conclusion maybe is not unexpected for V. Putin and D. Medvedev, but it is very interesting to know that if we are talking about Yeltsin's period of political life in Russia. It means that in 1994 the construction of new political system was not accompanied with traditional for democratic regimes political rhetoric.

Second conclusion is about R as a tool for text mining and graphics. This study shows that R is a working tool in political discourse studies. It can bring additional knowledge and illustrate results of scientific research. R has proved to be a valuable asset for quantitative research. However, to be completely honest we should say that there is a little limitation for Russian language in text mining with R. If one looks carefully in picture 1, he should notice that several one-rooted words are repeated with different ends. This error occurs because of limited support for Russian in NLP-package of R.

Литература

Балахонская Л. В., Быков И. А. (2018). Речевая агрессия в политических блогах радиостанции «Эхо Москвы» // Вестник Санкт-Петербургского Университета. Язык и литература. – Т. 15. Вып. 3. С. 492-506. DOI: https://doi.org/10.21638/spbu09.2018.313

Васильев А. Д. (2015). Лексико-фразеологические представления своего и чужого в Посланиях В.В. Путина Федеральному собранию (2012–2014 гг.) // Политическая лингвистика. – № 52. – С. 17–24.

Гавра Д. П. (2016). Послание как социальный феномен и особый жанр PR-текстов // Стратегические коммуникации в бизнесе и политике. – Т. 2. № 2. – С. 161–182.

Гаврилова М. В. (2017). Лингвистический анализ выступлений главы государства: тематика, направления и методы исследования // Политическая наука. – 2017. № 2. – С. 54-72.

Гаврилова М. В. (2013). Послание Федеральному Собранию // Дискурс-Пи. – 2013. № 3. – С. 111.

Косов В. В. (2019). Дискурсивные стратегии и тактики президентов России и Франции в формате прямого общения с аудиторией: модели коммуникации для управления конфликтными ситуациями // Журнал политических исследований. – №. 3. – С. 58-68.

Управляемость и дискурс виртуальных сообществ в условиях политики постправды / Под ред. Д. С. Мартьянова. – СПб.: ЭлекСис, 2019. – 312 с.

Brown G., Yule G. (1983). Discourse Analysis. – Cambridge: Cambridge University Press, 288 p.

Bykov I. A., Balakhonskaya L. V., Gladchenko I. A., Balakhonsky, V. V. (2018). Verbal aggression as a communication strategy in digital society. In: Proceedings of the 2018 IEEE Communication Strategies in Digital Society Workshop. St Petersburg. P. 12-14. DOI: 10.1109/COMSDS.2018.8354954

Dijk T. van. (2008). Discourse and Power. Contributions to Critical Discourse Studies. Houndsmills: Palgrave MacMillan, 320 p.

Fairclough N., Cortese G., Ardizzone P. (eds.) (2007). Discourse and Contemporary Social Change. Pieterlen: Peter Lang, 555 p.

Feinerer I., Hornik K., Meyer D. (2008). Text mining infrastructure in R. Journal of Statistical Software. Vol. 25(5). P. 1-54.

Feldman R., Sanger J. (2006). The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press, 424 p.

Hotho A., Nürnberger A., Paaß, G. (2005). A brief survey of text mining. LDV Forum. Vol. 20(1). P. 19–62.

Kabacoff R. I. (2011). R in Action! Data Analysis and Graphics with R. Shelter Island: Manning, 2011. 474 p.

Nosov A. V. (2018). Statistical analysis of near-synonymous words list and catalog in R. Vestnik of Saint Petersburg University. Language and Literature. Vol. 15, issue 3. pp. 453–464. https://doi.org/10.21638/spbu09.2018.310

Patocka-Sigłowy U. (2015). The Mechanism of Political Speeches Using the Example of the Address of President Vladimir Putin. Russian Linguistic Bulletin. 2015. Vol. 3(3). P. 24–27.

Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Ed. by G. Miner, D. Delen, E. Charlottesville, A. Fast, T. Hill, R. Nisbet. Amsterdam: Academic Press, 2012. 1000 p.

Schiffrin D. (1994). Approaches to Discourse. Malden: Blackwell, 482 p.

Stubbs M. (1983). Discourse Analysis: The Sociolinguistic Analysis of Natural Language. Oxford: Blackwell, 272 p.

Wodak R. (1989). Language, Power and Ideology: Studies in Political Discourse. – London: John Benjamins Publishing Company, 288 p.

References

Balahonskaja L. V., Bykov I. A. (2018). Rechevaja agressija v politicheskih blogah radiostancii «Echo Moskvy» [Verbal aggression in political blogs: A case of the "Echo of Moscow"]. Vestnik Sankt-Peterburgskogo Universiteta. Jazyk i literatura [Herald of St Petersburg University. Languages and literature]. Vol. 15. Issue 3. P. 492-506. – DOI: https://doi.org/10.21638/spbu09.2018.313 [in Russian]

Vasilev A. D. (2015). Leksiko-frazeologicheskie predstavlenija svoego i chuzhogo v Poslanijah V. V. Putina Federal'nomu sobraniju (2012–2014 gg.) [Lexical and phraseological representations of one's own and another's in the Messages of V.V. Putin to the Federal Assembly (2012–2014). Politicheskaja lingvistika [Political Linguistics]. № 52. P. 17–24. [in Russian]

Gavra D. P. (2016). Poslanie kak social'nyj fenomen i osobyj zhanr PR-tekstov [Address as a social phenomenon and a special genre of PR texts]. Strategicheskie kommunikacii v biznese i politike [Strategic communications in business and politics]. Vol. 2. № 2. P. 161–182. [in Russian]

Gavrilova M. V. (2017). Lingvisticheskij analiz vystuplenij glavy gosudarstva: tematika, napravlenija i metody issledovanija [Linguistic analysis of speeches of the head of state: topics, directions and research methods]. Politicheskaja nauka [Political Science]. 2017. № 2. P. 54-72. [in Russian]

Gavrilova M. V. (2013). Poslanie Federal'nomu Sobraniju [Address to the Federal Council]. Discurs-Pi [Discourse P]. № 3. P. C. 111. [in Russian]

Kosov V. V. (2019). Diskursivnye strategii i taktiki prezidentov Rossii i Francii v formate prjamogo obshhenija s auditoriej: modeli kommunikacii dlja upravlenija konfliktnymi situacijami [Discursive strategies and tactics of presidents of Russia and France in the format of direct communication with the audience: communication models for managing conflict situations]. Zhurnal politicheskih issledovanij [Journal of Political Studies]. №. 3. P. 58-68. [in Russian]

Upravljaemost' i diskurs virtual'nyh soobshhestv v uslovijah politiki postpravdy [Virtual Communities Manageability and Discourse in Post-Truth Politics]. Ed. By Martyanov. St. Petersburg.: ElekSis, 312 p. [in Russian]

Brown G., Yule G. (1983). Discourse Analysis. Cambridge: Cambridge University Press. 288 p.

Bykov I. A., Balakhonskaya L. V., Gladchenko I. A., Balakhonsky, V. V. (2018). Verbal aggression as a communication strategy in digital society. In: Proceedings of the 2018 IEEE Communication Strategies in Digital Society Workshop. St Petersburg. P. 12-14. DOI: 10.1109/COMSDS.2018.8354954

Dijk T. van. (2008). Discourse and Power. Contributions to Critical Discourse Studies. Houndsmills: Palgrave MacMillan, 320 p.

Fairclough N., Cortese G., Ardizzone P. (eds.) (2007). Discourse and Contemporary Social Change. Pieterlen: Peter Lang, 555 p.

Feinerer I., Hornik K., Meyer D. (2008). Text mining infrastructure in R. Journal of Statistical Software. Vol. 25(5). P. 1-54.

Feldman R., Sanger J. (2006). The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press, 424 p.

Hotho A., Nürnberger A., Paaß, G. (2005). A brief survey of text mining. LDV Forum. Vol. 20(1). P. 19–62.

Kabacoff R. I. (2011). R in Action! Data Analysis and Graphics with R. Shelter Island: Manning, 2011. 474 p.

Nosov A. V. (2018). Statistical analysis of near-synonymous words list and catalog in R. Vestnik of Saint Petersburg University. Language and Literature. Vol. 15, issue 3. pp. 453–464. https://doi.org/10.21638/spbu09.2018.310

Patocka-Sigłowy U. (2015). The Mechanism of Political Speeches Using the Example of the Address of President Vladimir Putin. Russian Linguistic Bulletin. 2015. Vol. 3(3). P. 24–27.

Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications . Ed. by G. Miner, D. Delen, E. Charlottesville, A. Fast, T. Hill, R. Nisbet. Amsterdam: Academic Press, 2012. 1000 p.

Schiffrin D. (1994). Approaches to Discourse. Malden: Blackwell, 482 p.

Stubbs M. (1983). Discourse Analysis: The Sociolinguistic Analysis of Natural Language. Oxford: Blackwell, 272 p.

Wodak R. (1989). Language, Power and Ideology: Studies in Political Discourse. – London: John Benjamins Publishing Company, 288 p.