European Proceedings of
Social and Behavioural Sciences
EpSBS

www.europeanproceedings.com
e-ISSN: 2357-1330

DOI: 10.15405/epsbs.2021.12.108

LATIP 2021
International conference «Language and Technology in the Interdisciplinary Paradigm»

CORPUS-BASED STUDY OF THE PERCEPTION OF RUSSIA IN THE AMERICAN DISCOURSE (1991-2021)

Elena Pozdnyakova (a), Ekaterina Suvorina (b)*
*Corresponding author

(a) Moscow State Institute of International Relations (MGIMO) 76, Prospect Vernadskogo, Moscow, 119454, Russia, helenpozdnyakova@yandex.ru
(b) Moscow City University, 4 Vtoroy Selskohoziajsvenny proezd, Moscow, 129226, Russia, SuvorinaEV@mgpu.ru

Abstract

We present the results of a corpus-based study of perception of Russia in the American discourse in the period of 1991-2021. The intrinsic cognitive processes are attributed to the shift in the lexical profiles, which have been elaborated using two corpora (News on the Web (NOW) and Corpus of Contemporary American English (COCA)). The period under analysis was divided into the subsets that cover identifiable historical periods before 2017. The years 2017-2021 were analysed separately by sectioning the corpora in order to study the fine structure of the temporal dynamics. The lexical profiles were elaborated using frequency data for the first left collocate in the pattern *adjective+Russia*. It is the most straightforward estimator of the hidden emotional state in the American discourse which, in turn, is driven by the cognitive processes stimulated by the external factors. The collocates are divided into four groups: neutral, historical, positive, and negative. The positive perception was stable in 1991-2020 and dramatically changed from January till April 2021. The negative perception was absent in 1991-2014 and in 2019. The negative emotional attitude raised since 2014 with an extreme growth in 2021. The nonstationary nature of the perception of Russia in the American discourse is proved by the statistical analysis.

Keywords: Corpus linguistics, collocation, collocates, Russia, statistical analysis, Web discourse

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

Contemporary cognitive linguistics addresses many aspects of our world. The special interest is attracted by its usage in studies of various emotional states. Cognitive linguistics utilizes the techniques of corpus linguistics. The latter provides reproducible results and places a strong mathematical and statistical background to the logical prepositions of a cognitive study.

The usage of corpora in cognitive studies is well described in (Brezina, 2018; Hoffmann et al., 2008; Stefanowitsch, 2020; Wallis, 2021). Nowadays, corpus managers provide plenty of statistical methods and a strong way to verify any hypothesis (Arppe, 2009; Cameron & Panović, 2014; Cohen et al., 2014; Flowerdew, 2012; Glynn & Fisher, 2010; Glynn, 2010; Haider, 2019).

The main assumption is that language usage is not a random process. Instead, it reflects the cognitive atmosphere in which a human being is located. Thus, by studying the corpus, we can address the estimation of the aforementioned cognitive atmosphere or, at least, we can identify the changes. The last item is very important, as far as we still have no proved quantities that can measure a cognitive state. Thus, the absolute measurement is impossible, but the relative shifts can be found out and interpreted manually.

The present situation in the foreign affairs between the East and the West, namely, between Russia and Western countries with the US leadership, is very tense. Somebody in mass media already compares the present day to the Caribbean crisis (1962) or the Able Archer crisis in Europe (1983). It is very hard to identify the difference between the actual propaganda and the true shift of the cognitive atmosphere in the opposite societies.

The linguistic study of the aggression markers has a long history (Balfour, 2019; Cortes et al., 2005; Cohen et al., 2014; Fernandez et al., 2009; Matsumoto & Hyisung, 2012), where three subtopics can be identified. The first item is concentrated on the “verbal dehumanization of objects of hatred or aggression” (Cortes et al., 2005). The second way is the study of cognitive complexity of speech, proving the rapid decrease of this complexity with conflict or crisis in progress. The third direction studies differences in the language used by political leaders.

The aforementioned studies mainly used brutal force attempts in sense of the direct calculation from a manually derived dictionary. Another approach consists of finding the special markers from a shortlist that surely presents an aggressive state.

We think that this is rather an artificial approach because of its nonuniqueness. Indeed, many dictionaries can be proposed and the choice of the texts for the analysis is also a matter of chance.

That is why we consider the cognitive aspects of the crisis under progress within the cognitive-corpus paradigm. Below we will concern only one small but clear factor, namely the perception. The study will use the power of the modern corpora of the American language (NOW, COCA) that covers the modern discourse in all aspects.

2. Problem Statement

We investigate the offset of the perception in the description of or mentioning Russia in American English. The emotional state is the simplest marker of the cognitive atmosphere in society. Moreover, it is

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

the easiest one to derive from the corpus using a collocation analysis, as far as it is based on the adjective usage analysis. Comparing to the cognitive complexity (Matsumoto & Hyisung, 2012) it does not mix all the words and it is not limited to manually selected documents or artificial dictionaries.

The objectivity of the study requires the homogeneous sampling, so that age, gender, nationality, etc., do not impact the lexemes distribution. Manual selection of the texts related to the crisis surely violates the homogeneous sampling principle as far as the authorship of white highly educated male politics is evident.

The usage of the corpora guarantees the maximum available homogeneous sampling across American society.

3. Research Questions

We want to investigate the shift of the perception in the description of Russia in American English. The emotional state is the straightforward marker of the cognitive atmosphere in society. Moreover, it is the easiest one to derive from the corpus using a collocation analysis, as far as it is based on the adjective usage analysis. Comparing to the cognitive complexity (Matsumoto & Hyisung, 2012) it does not mix all words, it is not limited to manually selected documents or artificial dictionaries. The objectivity of the study requires the homogeneous sampling, so that age, gender, nationality, etc., do not impact the lexemes distribution. The usage of the corpora guarantees the maximum available homogeneous sampling across American society.

The outline of the problem under study is as follows. First, we will identify the set of adjectives that are attributed to Russia using the corpora and their corpus managers. Next, we manually classify the obtained sets of words into historical, neutral, positive, and negative ones. This is not an arbitrary process, as far as it includes the well-known dictionary definitions. The offset of the perception of the adjectives will prove the actual change of the cognitive atmosphere affecting American modern society.

4. Purpose of the Study

The purpose of our study is the detection of the perception of Russia in the American discourse. We study this process in the diachrony covering the period from 1991 to 2021. Within these 30 years, we tagged manually the following subperiods:

1991-1993 - the appearance of modern Russia;

1994-1998 - Boris Eltsyn's presidential term;

1999-2008 - the first presidential term of Vladimir Putin;

2008-2013 - the world economic crisis;

2014-2016 - the active stage of the Ukrainian crisis.

The next period (2017-2021) is studied separately to refine the temporal dynamics.

Surely, the standard political or social approach creates a specific sub-corpus from mass media or politician's speeches and investigate manually the perception of Russia within them.

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

But in this way, we inevitably fall into the trap of the political discourse. Namely, political tendencies and rivalry among various politicians must be included. Thus, the aforementioned corpus would be extremely biased, and their homogenization becomes a very hard problem.

As far as we do not deal with politics but with the cognitive-emotional estimation from a natural language corpus, we decided to proceed with general corpora. They, indeed, include politicians and mass media. But they also include a tremendous amount of the web pages and other sources. Thus, they provide a reliable representation of the American discourse and our study is taking on the attributes of the general cognitive study.

We hypothesize that the perception is represented by some adjectives that collocate with the target word, namely, *Russia*. The world's perception of Russia changed in time. Some views were neutral, while some views were emotionally inflected. In some sense, it looks like a colour attributed to Russia. We try to find out the change of that colour in time and identify the underlying cognitive-emotional shift in the American discourse.

5. Research Methods

As far as our target is the studying of the diachronic change of the adjectives that collocate with the word *Russia* we use contemporary computer corpora. From the list of the available resources, we selected NOW corpus as well as Corpus of Contemporary American English (COCA).

The NOW corpus (News on the Web) according to the corpus home page (https://www.english-corpora.org/now/) "contains 12.4 billion words of the data from web-based newspapers and magazines from 2010 to the present time (the most recent day is 2021-04-21). More importantly, the corpus grows by about 180-200 million words of data each month (from about 300,000 new articles), or about two billion words each year".

The Corpus of Contemporary American English (COCA) is the only large, genre-balanced corpus of the American English. "COCA is probably the most widely-used corpus of English, and it is related to many other corpora of English. These corpora were formerly known as the "BYU Corpora", and they offer an unparalleled insight into the variation in English. The corpus contains more than one billion words of text (25+ million words each year 1990-2019) from eight genres: spoken, fiction, popular magazines, newspapers, academic texts, and (with the update in March 2020): TV and Movies subtitles, blogs, and other web pages" (https://www.english-corpora.org/coca/).

The NOW corpus was used for the years 2014-2021 while the COCA corpus was used for the years 1991-2013.

The request to the corpora was built in the form of LIST search in the form "Adjective + Russia". The following set of the corpus manager options was used: SECTION - target year (years), SORT - frequency.

Though it was mentioned earlier in (Hoffmann et al., 2008) that a collocation study becomes more reliable if more complex statistical measures are used (like log-likelihood or mutual information criteria), we limit ourselves to raw frequency data in this exploration analysis. The more in-depth study will use complex criteria if necessary. Practically, we obtain lexical profiles (Laufer, 2005) based on raw frequency data.

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

From the long lists, we artificially limit our profiles by the first 25 lexemes. The next step was the cleaning of the data from geographical terms, like Western Russia, Eastern Russia, etc.

The final lists with varying cardinality were analysed manually. Each member was attributed to one of four classes. The first class denotes negative connotation, the second one denotes positive connotation. The third class relates to the history and related issues. The last class relates to a so-called general description that doesn't indicate any perception. Finally, the number of each class member was calculated and stored by years.

The lexical profile reflects the in-depth cognitive patterns, so its change in time proves the cognitive shift. This is the shift of emotions that are stimulated by mentioning Russia.

6. Findings

The lexical profiles are presented in Table 01. This is a two-way table with both periods and positive/negative collocated adjectives. We removed neutral and historical collocations from the Table. The historical collocations were: *tsarist, Stalinist, bolshevik, post-communist, imperial, post-soviet, communist, soviet,* and *post-soviet*.

The neutral collocations were: *native, real, present-day, orthodox, rural, old, neighbouring, nuclear-armed, veto-wielding*. We consider the last list as neutral as far as they appear in the informational context.

The trends of the total count for positive and negative collocations are shown in Figure 01. It is evident that positive perception was stable and almost stationary for 30 years. The dramatic jerk appears only in 2021. Recalling history, we see that neither the Chechen wars nor the Ukrainian conflict, as well as Donald Trump election case, had not been changing the level of the positive perception of Russia in the American English language. Thus, we might make the conclusion that widely reported a veritable outcry in the press was nothing but a kind of propaganda, stimulated and valued by some rather small political groups. In the large group of Americans, all the aforementioned historical events did not modify the cognitive frames which are responsible for the positive value of Russia.

Regarding evidence from Figure 01 it is worthwhile to note that an additional statistical check was evaluated against the null hypothesis of the uniform distribution.

The test against the uniform distribution was evaluated employing chi-squared Pearson critical value. We use the online statistical calculator (https://math.semestr.ru/group/hypothesis-testing.php). For the negative connotations observed statistics is 43.28 against critical value 14.1. For positive connotations observed statistics is 111.5 with the same critical value.

The observed Pearson value satisfies the condition "observed" > "critical" thus we reject the null hypothesis and conclude that the distribution is nonuniform in both cases. As far as the variation coefficient is less than 30% then both populations are homogeneous, and results are reliable.

Negative perception varied significantly within 30 years. Before 2014 there is no evidence of the negative perception of Russia at the top of the lexical profile for the adjective collocates.

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

Table 1. Corpus-based lexical profiles of the word "*Russia"* (1991-2021)

Period	Positive	Negative
1991-1993	democratic	disintegrating
	free	
	independent	
	great	
1994-1998	democratic	
	resurgent	
	cooperative	
	better	
	beloved	
1999-2008	resurgent	
	contemporary	
	democratic	
	open	
2008-2013	resurgent	
	open	
	democratic	
	holy	
2014-2016	resurgent	aggressive
	great	assertive
	encouraging	
	greater	
2017	ongoing	encroaching
	new	resentful
	strong	tangled
	resurgent	sweeping
	renowned	so-called
		sensationalized
		punishing
2018	ongoing	weak
	top	Trump
	resurgent	rigged
	strong	fricking
	open	fake
2019	democratic	
	modern	
	contemporary	
	resurgent	
2020	resurgent	imperial
	open	assertive
	modern	aggressive
	ongoing	
	top	
	resurgent	regime-backer
2021		revanchist
		doping-tainted
		sanctions-hit
		assertive
		unwashed
		expansionist
		phony

https://doi.org/10.15405/epsbs.2021.12.108
Corresponding Author: Ekaterina Suvorina
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

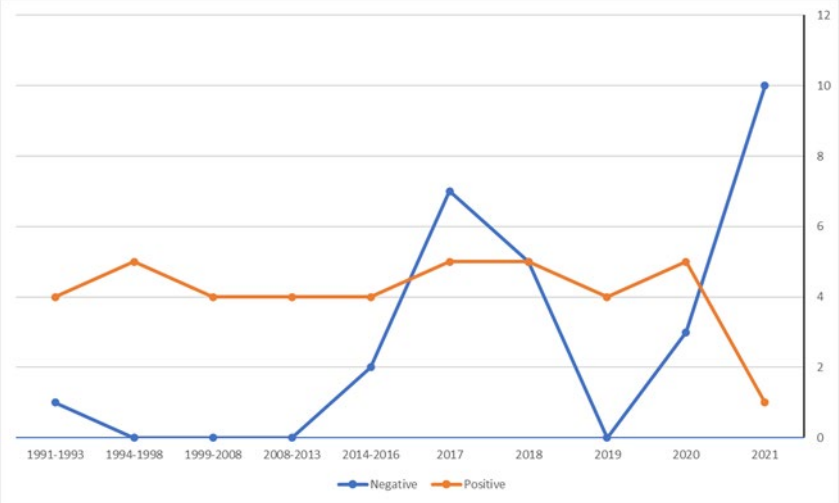

Figure 1. Evolution of positive and negative collocations in the lexical profiles of the word *Russia*

The first local maximum is observed in 2017. The rise began in 2014. The gravest collocates were observed in 2018. This process abruptly stopped in 2019 with an extreme growth in 2021. The last lexical profile (for the first 4 months of 2021) covers the negative perception for the international politics as well as for sport. Moreover, in 2021 negative perception that had a plaque of dehumanization appeared. Namely, collocates *unwashed* and *phony*.

7. Conclusion

We have evaluated a corpus-based study of the perception of Russia in the American discourse in 1991-2021 years. The lexical profiles have been elaborated using two corpora (NEW and COCA) for the pattern *adjective+Russia*. The positive perception was stable and stationary for 1991-2020. The negative perception was negligible before 2014 with a further rise. The disappearance of negativism in 2019 is unexplained but, in turn, it switched to the extreme growth of the negative perception together with the fall of positivism in the first three months of 2021.

Acknowledgments

We acknowledge the web resource http://english-corpora.org for the excellent corpus manager

References

Arppe, A. (2009). Linguistic choices vs. probabilities – how much and what can linguistic theory explain? *The Fruits of Empirical Linguistics*, *1*, 1-24. https://doi.org/10.1515/9783110216141.fm

Balfour, J. (2019). 'The mythological marauding violent schizophrenic': Using the word sketch tool to examine representations of schizophrenic people as violent in the British press. *Journal of Corpora and Discourse Studies*, *2*, 40-64. http://doi.org/10.18573/jcads.10

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316410899

Cameron, D., & Panović, I. (2014). Corpus-based discourse analysis. *Working with written discourse*, 81-96. https://www.doi.org/10.4135/9781473921917

Cohen, K., Johansson, F., Kaati, L., & Mork, J. C. (2014) Detecting Linguistic Markers for Radical Violence. *Social Media, Terrorism and Political Violence*, *26*(1), 246-256. https://doi.org/10.1080/09546553.2014.849948

Cortes, B. P., Demoulin, S., Rodriguez, R. T., Rodrigues, A. P., & Leyens, J.-P. (2005). Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. *Personality and Social Psychology Bulletin*, *31,* 243-253.

Fernandez, I., Paez, D., & Pennebaker, J. W. (2009). Comparison of expressive writing after the terrorist attacks of September 11th and March 11th. *International Journal of Clinical Health Psychology, 9,* 89-103.

Flowerdew, L. (2012). How is Corpus Linguistics Related to Discourse Analysis? *Corpora and Language Education*, 81-110. https://doi.org/10.1057/9780230355569_4

Glynn, D. (2010). Synonymy, Lexical Fields, and Grammatical Constructions. A study in usage-based Cognitive Semantics. In: H.-J. Schmid, & S. Handl (Eds), *Cognitive Foundations of Linguistic Usage-Patterns* (pp. 89-118). https://portal.research.lu.se/portal/files/5764004/1647480

Glynn, D., & Fisher, K. (2010). *Quantitative methods in cognitive semantics: corpus-driven approaches.* https://doi.org/10.1515/9783110226423

Haider, A. S. (2019). Using corpus linguistic techniques in (critical) discourse studies reduce but does not remove bias: evidence from an Arabic corpus about refugees. *Poznań Studies in Contemporary Linguistics, 55*(1), 89-133. https://doi.org/10.1515/psicl-2019-0004

Hoffmann, S., Evert, S., Smith N., Lee, D. Y. W., & Berglund, Y. (2008). *Corpus linguistics with BNC-web -A practical guide*. Frankfurt am Main: Peter Lang.

Laufer, B. (2005). Lexical Frequency Profiles: From Monte Carlo to the Real World. A response to Meara. *Applied Linguistics, 26*(4), 582-588. https://doi.org/10.1093/applin/ami029

Matsumoto, D., & Hyisung C. Hwang (2012). The Language of Political Aggression. *Journal of Language and Social Psychology, 32*(3), 335-348. https://doi.org/10.1177/0261927X12460666

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

Wallis, S. (2021). *Statistics in Corpus Linguistics Research - A New Approach*. Routledge.