

Russia is not Putin: Collocational analyses of *Russia*, *Russian(s)*, and *Putin* in US Mass Media

Ulbossyn Parmanova

University of Georgia

ulbossyn.parmanova@uga.edu

Abstract

This paper investigates the collocational network around the words *Russia*, *Russian(s)*, and *Putin* in news text. Association analyses show that both *Russia* and *Russian(s)* are closely associated with warfare words. Interestingly, *Putin*, the third word studied, is associated with Ukraine rather than Russia.

1 Introduction

In the context of the ongoing war, as well as historical events linked to Russia-USA relations, the words *Russia*, *Russians*, and *Putin* have gained stereotypical interpretations and uncertain meanings. Using association measures and keyword analyses, this paper performs a corpus analysis of the connotative meanings of *Russia*, *Russian(s)*, and *Putin*, looking at their representation in the US press between 1860-2010.

2 Previous Work

A review of existing work reveals a predominantly negative representation of Russia in Western mass media. Indeed, [Repina et al. \(2018\)](#) conclude that “The image of Russia in Western media is negative,” and that Russia has “the image of the enemy,” (p. 562). Similarly, [Ilyushkina & Chudinov \(2019\)](#) state that the metaphorical model of Russia as “enemy” in American press is based upon “the negative image of the Russian leader, the negative characteristics of Russia, [and] the negative characterization of relations between Russia and the United States,” (p. 28). There is also research suggesting an increase in negative perception of Russia in articles in the New York Times following the annexation of Crimea in 2014 ([Smirnova et al., 2017](#)).

3 Methods and Corpora

3.1 Corpus

This research examines *Russia*, *Russian(s)*, and *Putin* in news corpora using collocations. Attestations come from the newspaper subcorpus of the Corpus of Historical American English (COHA). The subcorpus is comprised of 43,507,224 words, consists of material between 1861-2007, and includes such newspapers as the New York Times, Chicago Tribune, Wall Street Journal, Christian Science Monitor, Boston Globe, USA Today, Washington Post, Associated Press, San Francisco Chronicle, Houston Chronicle, Atlanta Journal Constitution, Denver Post, and CSMonitor.

3.2 Metric

The association measure used here is Log-Likelihood (LL). LL is a combined metric that measures both frequency and content-related words ([Baker 2006](#), p. 102). It calculates both function and content words in a span. [Brezina \(2018\)](#) states that the LL metrics highlights “two main dimensions: frequency and exclusivity (words cooccur only/predominantly in each other’s context defined by the span) of the collocational relationship,” (pp. 71-74).

4 Data Analysis

Data were analyzed using the PolmineR package ([Blaette, 2020](#)) in RStudio.

Aiming to find the most important collocates and to build a collocation network for the selected lexical items, the words that co-occur with *Russia*, *Russian(s)*, and *Putin* in a five-word window size were identified. Following a search for word collocations, a keyword procedure was carried out

comparing attestations across corpora. Keyword techniques are helpful in identifying the key concepts related to selected tokens and can be used to compare their lexical development diachronically. A random co-occurrence baseline provides the corrected values of expected frequency of the collocations.

5 Results

Querying for *Russia* retrieved 5,206 hits within the 30-word span to the left and the right in the NEWS subset of COHA and revealed two major groups: one related to Russian regime (*soviet, Russian, Stalin, satellites, Czar, Lenin, revolution, Bolshevism, trade*) and one related to war and international relations (*war, German, Hitler, treaty, peace, attack, ally, invasion*).

Metrics for the occurrences of the *Russian(s)* collocates were also mainly related to “wartime words,” mirroring the findings for the node *Russia*. Attestations of *Russian(s)* in COHA by year illustrate increased Western news discussion of Russians in 1904, during World War I, and at the start of World War II. Notably, reports about *Russians* were still significant during 1950-1990, presumably because of post-war relations and Cold War tension after WWII. The collocational link between *Russia* and *Russian(s)* was strong, being ranked high (LL score = 182.5, rank_ll = 18).

Compared to the previous findings, there was a very small sample collected for the node *Putin*, with a total of 31 attestations. The first mention of *Putin* in the COHA newspapers is dated 2000, the year he became president. The list of 30 closest associates of *Putin* comprises such proper names as *Bush, Assad, Cuba, Kremlin, Mikhail Gorbachev, Rice*, etc. More interestingly, *Ukraine* (but not *Russia*) appeared to be in the list of significant collocates of *Putin*. Further, the attestation of *Putin* in relation to *Russia* ranked low (LL=4.4), which speaks to the fact that Putin is not representative of Russia in the 2000s.

6 Conclusion

In conclusion, querying COHA for both *Russia* and *Russian(s)* retrieved collocates mainly related to the lexicosemantic group “war.” Although *Russia*, *Russian(s)*, and *Putin* are nodes within a network of related lexical items, there was no strong relationship found between the words *Putin* and *Russia*.

Limitations

Results from corpus-based analyses are limited and biased by the corpora considered. The results presented here are biased in that they are derived from news text of American English. It is conceivable that corpora of news texts in other languages and from other countries could lead to different results.

Ethics Statement

The analyses performed here pose no significant risk of harm.

Acknowledgment

Acknowledgement of computing resources is given to the University of Georgia Advanced Computing Resource Center. Thanks is given to John Hale, Chad Howe, and Donald Dunagan for their valuable feedback.

References

- Anastasia Smirnova, Helena Laranetto, and Nicholas Kolenda. 2017. *Ideology through sentiment analysis: A changing perspective on Russia and Islam in NYT. Discourse & Communication*, 11(3) 296–313. <https://journals.sagepub.com/doi/pdf/10.1177/1750481317699347>.
- Andreas Blaette. 2020. Package “polmineR”. <https://github.com/PolMine/polmineR>.
- Ekaterina A. Repina, Marina R. Zheltukhina, Natalya A. Kovaleva, Tatiana G. Popova, and Conchita Garcia Caselles. 2018. *International media image of Russia: trends and patterns of perception. XLinguae European Scientific Journal*, 11(2):557-565. http://www.xlinguae.eu/files/XLinguae2_2018_45.pdf.
- Maria Yu. Ilyushkina and Anatoly Chudinov. 2019. *Metaphorical Modeling of Russia's Image: the case of American press. Yazyk I kul'tura*, 45:20-30. <https://doi.org/10.17223/19996195/45/2>.
- Paul Baker. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Stefan Th. Gries. 2022. *Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more. Research Methods in Applied Linguistics*, 1(1). <https://doi.org/10.1016/j.rmal.2021.100002>.
- Vaclav Brezina. 2018. *Statistics in Corpus Linguistics. A practical Guide*. Cambridge University Press, Cambridge, UK.