2021-07-30

Linguistics of Russian Media During the 2016 US Election: A Corpus-Based Study

Devon K. Terry
Brigham Young University

Linguistics of Russian Media During the 2016 US Election:

A Corpus-Based Study

Devon K. Terry

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Jeff Parker, Chair
Earl K. Brown
William G. Eggington

Department of Linguistics

Brigham Young University

ABSTRACT

Linguistics of Russian Media During the 2016 US Election:
A Corpus-Based Study

Devon K. Terry
Department of Linguistics, BYU
Master of Arts

The purpose of this study is to perform a linguistic analysis of Russian mass media focused on its coverage of the 2016 US presidential election. More specifically, it will be a corpus-based study, using the corpus as a foundational source for quantitative and qualitative data. This study will use a collection of keywords from the corpus and analyze their contexts as they pertain to the two primary candidates, Hillary Clinton and Donald Trump. This study uses corpus linguistic research tools such as sentence tokenization, Key Words in Context (KWIC), sentiment analysis, word embedding visualization, word-vector math, word frequency lists, and collocate analysis as part of the quantitative analysis. The results of the sentiment analysis and word vector analysis show a moderate bias in the corpus favoring Donald Trump. Additionally, a more in-depth qualitative analysis of sentences from the corpus containing the keywords is performed. A framework using Appraisal Theory is used to examine sample sentences to show how the corpus evaluates and appraises the candidates. The qualitative analysis shows how many sentences in the corpus are full of judgment towards Hillary Clinton, positive appraisal of Donald Trump, and attempts to expand a positive dialog surrounding Donald Trump, as opposed to a contraction of dialog and expansion of negativity about Hillary Clinton. The predicted Russian geopolitical agenda seeks to demean American politics, positively influence the perception of Russians towards Vladimir Putin, as well as support Donald Trump insofar as his policies align with Russia's goals.

Keywords: corpus, russian, linguistics, sentiment, embeddings, media, election, 2016, Hillary Clinton, Donald Trump

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: Russian News Media and Political Agendas

1.1 - Political Discourse and Corpus Linguistics

Corpus linguistics studies are useful for their empirical and data-driven methodologies. They are especially adept at discovering general patterns and trends in context for specific words and topics due to the utilization of large corpora. Studies utilizing corpus linguistics methods typically focus on specific words or topics, analyzing "How X is talked about." The individual words symbolize the central hub in a network of analysis, where the nearby context and broader neighborhood carry the meaning and political message. Ultimately, according to Ädel, these studies are interested in discourse patterns, argumentation patterns, or cultural beliefs (Ädel 2010).

1.2 - Media and Russian Politics

In contrast to the freedom of press enjoyed in many other countries, the national and regional governments of Russia control or, at the very least, exert great influence on news media circulated in Russia (Rozenas & Stukal 2019; Szostek & Hutchings 2015). And despite overt intrusion of the state government into the dissemination of information, several surveys have found most of the population of Russia believe state-run television networks are the most unbiased and most reliable sources of information (White, Oates & McAllister 2005). This inherent trust in government media allows the state to overtly set political agendas, mold public sentiment, and frame national and global events.

It is clear that Russia has an agenda when it comes to elections. Many studies have been performed about the relationship between the media and elections in Russia (White, Oates & McAllister 2005; Oates 2014; Darczewska 2014; Field et al. 2018; Rozenas & Stukal 2019). Specifically, White (2005) studied the effects of media on the 1999 Russian Parliamentary (Duma) and the 2000 Russian Presidential elections. Unsurprisingly, he found a significant correlation between viewing state-sponsored media and voter tendencies, all other factors being equal. Though it is unclear whether state-sponsored news media attracted like-minded voters, or whether it was the impetus for those voting tendencies, the research found a significant correlation in voting patterns that differed from previous elections.

The state-run media also employs agenda-setting and consistent narratives in order to shape perception of global events. Oates (2014) examined Russian state narratives in the media, specifically about the attack on the 2014 Malaysia Airlines Flight 17. She explains that "strategic state narrative is a useful proxy for the state's intentions and terms of engagement with the world. It cannot deviate too radically from the reality of the state, but at the same time it both signals intentions and aspirations on the world stage" (3). In other words, the facts will still be stated, but the interpretation of the events are key in shaping public opinion.

Additionally, Darczewska (2014) explores how Russia is interested in molding public sentiments in favor of a "Russian World." She explores the information warfare surrounding the annexation of Crimea from Ukraine. She explains that Russia, to retain its image and support of the people, disseminated propaganda and misleading information throughout Russia and Ukraine to shape public opinion. Specifically, she walks through the propogandist strategies Russia used in order to wage a defensive information war during the Crimean annexation. These strategies included emotional agitation, "supposed obviousness", simplification, and manipulated expectations (2014:25).

These are only a handful of examples of the research that has been conducted about more overt displays of political manipulation of and through the media. Additionally, more subtle methods of such manipulation have been employed and researched. Field (2018) explores how the subtle manipulation of what topics are covered in Russian media, in tandem with *how* aspects of those topics are presented, are used to set a political agenda without making overt statements. She explores how frequency of references to the US in a specific newspaper (Izvestya) are inversely correlated to the rise and fall of Russia's GDP (Field et al. 2018). In other words, the newspaper surreptitiously omits mention of the US when the economy is in a bull market. But, when the economy is in a bear market, mentions of the US overtly skyrocket, as if to blame the US for any negative economic trends.

This research is reinforced by Rozenas & Stukal (2019), who make the following claim about autocrats (using Vladimir Putin as an example):

> "[They] increasingly employ information manipulation to create a perception that they are competent economic managers" and that they are "not as concerned about whether certain facts are reported but [are] more concerned about how those facts will be interpreted. To impose a specific interpretation that benefits the government requires the media to frame facts in a certain way. If the media is successful at persuading citizens that external factors are culpable for economic failures while the domestic government is responsible for economic accomplishments, then the facts themselves—without the underlying interpretation—cannot directly hurt or benefit the government" (Rozenas & Stukal, 2019: 13).

1.3- Recent Narratives and the 2016 Election

Large-scale events of this past decade provide examples of specific Russian narratives that have been propagated in Russian domestic news media. The crisis in Ukraine and the annexation of Crimea are two examples of events that provide evidence for a consistent narrative and geopolitical agenda.

Concerning the Ukraine Crisis, Hutchings (2015) analyzes narratives that were played out on the news channel Rossiya 1. These narratives centered around anti-western sentiments. According to his research, the West, who were meddling in Ukraine, were hypocritical, arrogant, had 'Double Standards', and were guilty of war crimes. Accordingly, any accusations from such countries about Russia's involvement in Ukraine were not to be taken seriously (Szostek & Hutchings 2015). Additionally, the narrative surrounding Crimea and its importance to Russian geopolitics (at the expense of Ukraine and Crimeans who opposed annexation) has been reinforced for so long, that it has become an accepted fact among Russians that Crimea belongs to Russia, despite what Ukraine and the rest of the world think. Suslov (2014) analyzes how this ideology has infiltrated all levels of Russian media, from news outlets, to social networks, to the blogosphere. She explains: "A geopolitical master-narrative helps users distance themselves from morality and the law of nations when speaking about the annexation of Crimea. As one user argues, '[national] interest and force reign in geopolitics …' and in the same vein, 'geopolitics is beyond morality … [it's about] national interests, which are far from human interests'" (Suslov 2014: 598). This demonstrates how potent a 'Master-Narrative' or agenda can be for certain political issues.

Narratives surrounding these two events are relatively straightforward, and the research shows a definitive agenda that the Russian government wants to convey to its citizens. Less clear is the internal narrative Russia had during the 2016 US presidential election.

Many researchers have studied Russia's meddling in US media and social platforms prior to the November 2016 ballot (Boyd et al. 2018; Badawy, Ferrara & Lerman 2018; Padda 2020). From Twitter bot accounts to Facebook ads, to targeted misinformation campaigns, the general effects of Russian election interference can still be felt in the current US political environment. Russia had an external agenda on how it wanted US citizens to vote and feel towards specific issues and candidates. However, it is less certain what Russia's internal agenda towards its own citizens was. Kazun and Kazun (2017) explain how Russia drew the attention of the Russian people towards the election and explain how the media was partly responsible for Russia being the only country to favor Donald Trump among 45 countries polled. Similarly, Stokes (2017) provides evidence for overwhelmingly negative coverage of Hillary Clinton and the Democratic party in Russian mass media. However, the purpose and agenda for this are still hidden. And despite a preference for the Republican candidate, the coverage of the election was far from black and white. Both candidates were often portrayed in negative lights, and factual reporting of scandals on both sides were rampant. Burrett (2018) explains that a general negative outlook on American democracy is prevalent on Russian television. Additionally, she speculates about why a particular candidate may get positive and negative coverage, despite being favored: "Propaganda…works best when it is not absolute. Some balance and ambiguity in reporting provides the veneer of objectivity" (p. 319). My research aims to pierce that "veneer of objectivity" and discover the themes, agendas, and narratives that the Russian government hoped to convey to its people during the 2016 campaigns.

1.4 - Research Questions

This paper, through analysis of linguistic trends and patterns, will answer the following research questions using a Russian media corpus as a basis for the research:

R1) What sentiments were made towards the candidates of the 2016 US Election?

R2) How did the Russian media frame the candidates to Russian citizens?

R3) How do these trends reflect on possible Russian geopolitical agendas?

Chapter 2: Research Design and Methodology

2.1 – Corpus Design

In political discourse studies, politics are analyzed using either a narrow or broad definition. A narrow definition of political discourse is when political figures, whether in or out of government, communicate about political matters for political purposes (Ädel 2010). Contrastingly, a broad definition of political discourse might include any communication which happens to be on a political topic. To thoroughly answer the previously listed research questions, this study will use a broad definition of politics and will utilize both quantitative and qualitative methods.

To study the US election as portrayed by Russian Media on a broad scale, I used a relevant and appropriate corpus containing Russian news articles from the period surrounding the 2016 election. Specifically, I used the Russian News Media corpus designed by Max Formichev (https://github.com/maxoodf/russian_news_corpus). With his permission, the corpus was used to answer the study's research questions. The corpus contains about 1.5 million articles, over 360 million tokens, and over 5 million unique words. The time period of the collected articles ranges from April 2016 through March 2017, coinciding with the main timeline for the campaigning and election season. The corpus was provided in lemmatized form, meaning verbs are in the infinitive and adjective and nouns have been reduced to the nominative case. The text of the corpus was provided in a single text document without word or sentence tokenization and without tagging of linguistic features.

Below is a list of websites the articles in the corpus were drawn from:

http://echo.msk.ru	http://ria.ru	http://www.fontanka.ru	http://www.newizv.ru	http://www.rosbalt.ru
http://izvestia.ru	http://tass.ru	http://www.gazeta.ru	http://www.ng.ru	http://www.vedomosti.ru
http://lenta.ru	http://utro.ru	http://www.interfax.ru	http://www.novayagazeta.ru	http://www.vesti.ru
http://newsru.com	http://vz.ru	http://www.kommersant.ru	http://www.ntv.ru	
http://regnum.ru	http://www.1tv.ru	http://www.kp.ru	http://www.pravda.ru	
http://rg.ru	http://www.aif.ru	http://www.mk.ru	http://www.rbc.ru	

Due to the nature of Russian news media, as well as the unknown biases of the person who compiled the corpus and selected which news sites to draw from, this could introduce bias into the data presented in the corpus. However, due to the robust size of the corpus as well as the number of different news sites selected, the author is confident in the relative objectivity of the data.

2.2 – Quantitative Tools for Corpus Analysis

The corpus compiled from these various news sources was analyzed using Word2Vec, the Python module NLTK (Bird et al. 2009), and WordCruncher (Rosenvall & Shelley 1988) in order to produce frequency lists, tokenize sentences, create word embedding models, and keyword lists, as well as perform an automated sentiment analysis of sentences that contain keywords. Corpus-assisted political discourse studies often compare keywords and how they are used in other registers or corpora. In these types of studies, keywords can be data driven, meaning an algorithm or corpus software selects keywords based on frequency lists. Or, keywords can be pre-selected by the researcher when a specific word or topic is the focus of the study (Ädel 2010). The keywords in this study were selected using a collocate report in WordCruncher by isolating the top words related to US and Election that appeared most frequently in the corpus near those words. An analysis of these keywords helped unveil overarching themes and topics reflected in the corpus.

Once these keyword lists, vector models, and tokenized sentences were compiled and constructed, this quantitative data was analyzed in Chapter 3 to provide insight into the corpus as well as into each keyword. The quantitative data was also used to understand the relationships, sentiments, and general usage of keywords by the Russian media. An analysis of observed patterns in word frequencies, sentiment scores, embeddings, and keywords in contexts was used to analyze possible Russian geopolitical agendas.

Similar methods have been utilized in previous studies. For example, Kazun & Kazun (2017) used a type of word embedding in their own analysis of the 2016 election. They pre-selected 23 keywords which included candidates, political parties, ideological issues, and scandals. After running these keywords through the Word2Vec algorithm, they were able to visualize with a two-dimensional network graph the strength of the connections between keywords. This was based on how frequently the words were found together, and how far apart they were in context. They did this for articles from one month before the election compared to one month after the election. Part of their analysis involved the change in these connections over time, as well as the general strength of word-level connections.

The method of word vectors/embeddings will analyze the corpus holistically to find patterns in word-level connections and relationships. Similar analysis will be performed using the other methods as well. Neighborhood reports, collocates, and Key Words in Context (KWIC) will be used to identify relevant sentences which will be used for both the quantitative and qualitative textual analysis. And though the bulk of the qualitative analysis occurs in Chapter 4, in Chapter 3 a short qualitative analysis using some of examples follows each of the quantitative methods. The sentence examples were chosen by sorting the sentences by keyword, polarity, and strength scores. Once sorted, the sentences were chosen based on a qualitative assessment of their pertinence to the topic as well as by observing trends and patterns in thousands of sentences. The English translations following each sentence example are my own.

The final objective of the research is to discover recurring themes and patterns found in the corpus that reflect a Russian agenda and narrative towards the 2016 election.

2.3 – Qualitative Analysis using Appraisal Theory

In Chapter 4, a qualitative analysis will be provided. To better categorize sentences beyond basic sentiment polarity and strength scores or word vectors, I will utilize a framework for discourse analysis to guide the qualitative analysis. Chapter 4 uses this framework to answer

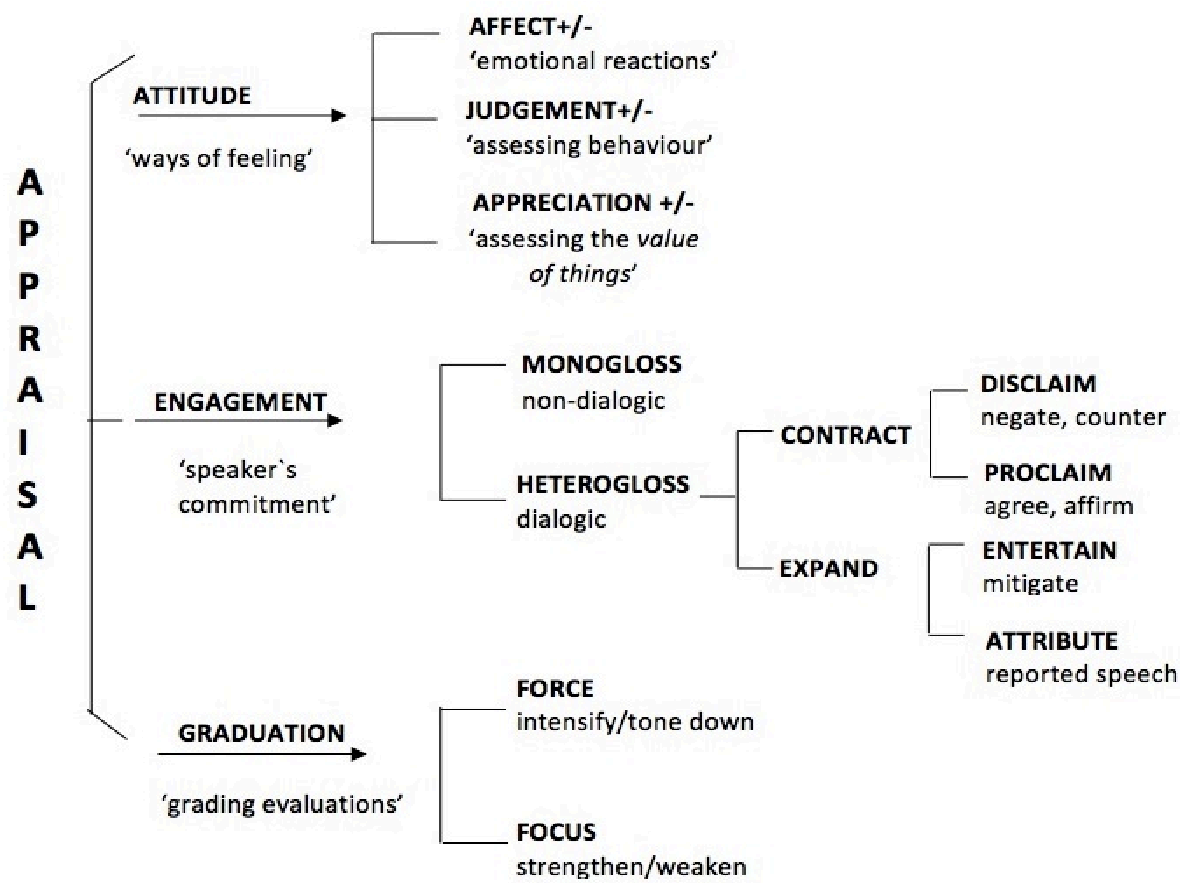

Figure 1. Appraisal Theory Framework (Martin and White, 2005, p. 38)

The remaining two research questions R2 and R3. This framework was applied to sentences from the corpus to show how the media engaged with the candidates, how they framed the election to its citizens, and what specific agenda might underly this framing. The framework is based on

Appraisal Theory from the work of Martin and White (2005), supplemented by Oteiza (2017). An outline of the framework is shown in *Figure 1.* Appraisal Theory includes a definition of appraisal actors (Appraiser and Object), and three main semantic areas: Graduation, Attitude, and Engagement. In Appraisal Theory, the actors are the "Appraiser" and the "Object of Appraisal". For the purposes of this study, the "Appraiser" is the Russian News Media, and the "Object of Appraisal" is the 2016 US Election. The keywords and sample sentences under observation were analyzed to discover the interactions and relationships between Appraiser and Object. Additionally, the target audience of the appraisal is the Russian society/people. The most common objects of appraisal in the corpus were Clinton and Trump. Other topics and events, like the cyberattack, various scandals, and US policy in general will be touched on briefly, but the keywords providing the most insight into the framing of the election are *Clinton* and *Trump.*

Chapter 3: Sentiment Analysis, Collocates, and Word Vectors

Research Questions:

R1) What sentiments were made towards the candidates of the 2016 US Election?

R2) How did the Russian media frame the candidates to Russian citizens?

R3) How do these trends reflect on possible Russian geopolitical agendas?

For this chapter analysis, a quantitative analysis was performed which will help inform the qualitative analysis in Chapter 4. This chapter will answer research question R1 using keywords, their collocates, sentiment values, and word vectors.

3.1 - Choosing Keywords

In order to obtain a relatively unbiased selection of keywords, the corpus was indexed using the WordCruncher software (Rosenvall & Shelley 1988). The software can generate a collocate report based on multiple keywords within specified vicinity of each other. Because the topic in question is the United States Election, the words *US* and *Election* were chosen, with an inclusive range of 15 words to the right and to the left. The top 70 collocates in the corpus, in areas where *US* and *Election* were found within 15 words of each other, were chosen. These 70 were further pared down to exclude function words, numbers, and other exemplars not directly related to the 2016 US Election. The remaining 25 keywords (with translations) are as follows:

сша	USA	*голосование*	vote
президент	president	*хакер*	hacker
трамп	Trump	*демократ*	democrat
клинтон	Clinton	*пересчет*	recount
хиллари	Hillary	*ноябрь*	November
дональд	Donald	*пенс*	Pence
республиканский	republican	*сандерс*	Sanders

демократический	democratic	*кампания*	campaign
кибератака	cyberattack	*гонка*	race
инаугурация	inauguration	*голосовать*	to vote
демпартия	democratic party	*теледебаты*	debates
кандидат	candidate	*бенгази*	Benghazi
партия	party		

3.2 – Automated Sentiment Analysis

An automated/computational sentiment analysis was performed on sentences containing

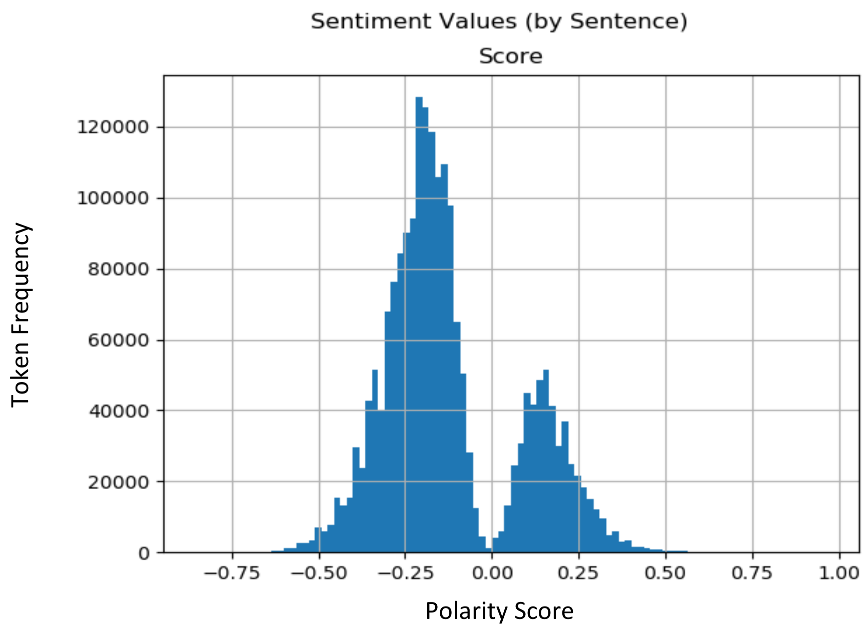

Figure 2. Sentiment Histogram for all Sentences

the keywords. The corpus was tokenized at the sentence level using the open-source Python module NLTK (Bird, Klein & Loper 2009). Upon tokenization, each sentence was checked for the presence or absence of a keyword. If the sentence contained a keyword, it was stored in a data frame, and categorized by keyword. Each keyword sentence was then processed with a sentiment analyzer. The sentiment analysis tool that was used was the Dostoevsky sentiment analysis python library (https://github.com/bureaucratic-labs/dostoevsky), trained on the RuSentiment dataset (Rogers et al. 2018). The results for each sentiment analysis included a polarity category (Positive, Negative, or Neutral) and a strength score of 0 to 1, with 0 being weak and 1 being strong. Sentences with neutral polarity were excluded from the dataset because in almost all cases the sentence only contained function words or were very short (i.e., 2-3 words long), providing little to no context for the keyword. Sentences with a positive or negative polarity were collected into a database and the collective data was visualized according to strength scores. Figure 1 shows a bi-modal distribution of sentiment values, with a sharp concentration of weak to moderately negative sentiments. When compared to the distribution of all sentences in the corpus (as opposed to these keyword sentences), a similar pattern emerges. The histogram for the

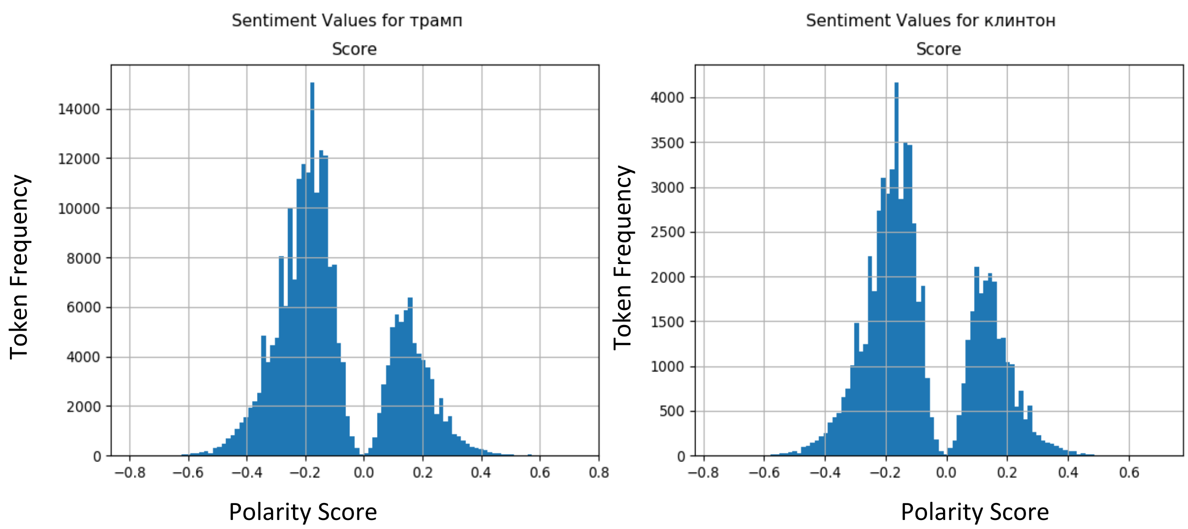

Figure 3. Sentiment Values for Trump (Left) and Clinton (Right)

entire corpus looks almost identical to the keyword histogram. Additionally, this bimodal pattern is consistent across histograms for each individual keyword (see *Trump* and *Clinton* above), with slight variation in degree of kurtosis and variance. This pattern could suggest that Russian news media in general shows moderate to low levels of sentiment or strength of emotion, with the majority of sentences reflecting negative polarity. And even when keywords do show some variation in their sentiment values, it is not a large variation. This could be due to the nature of news language, and could also reflect the nature of the Dostoevsky library (https://github.com/bureaucratic-labs/dostoevsky) and RuSentiment dataset (Rogers et al. 2018), which used social media posts as training data, which can often contain more extremes of emotion. However, the fact that there is variation in how sentences are categorized as well as variation in their strength scores leads to the assumption that news language is by nature less sentiment filled. In a study about news sentiment pertaining to economics, Shapiro, Sudhof, and Wilson found that human classifiers rated most news articles as "Neutral Sentiment" on a scale of 1 to 5, with 1 being extremely negative and 5 being extremely-positive (2020). It makes sense that an algorithm or model would similarly trend towards more neutral sentiment with more precise decimal measures, if not necessarily more accurate measurements.

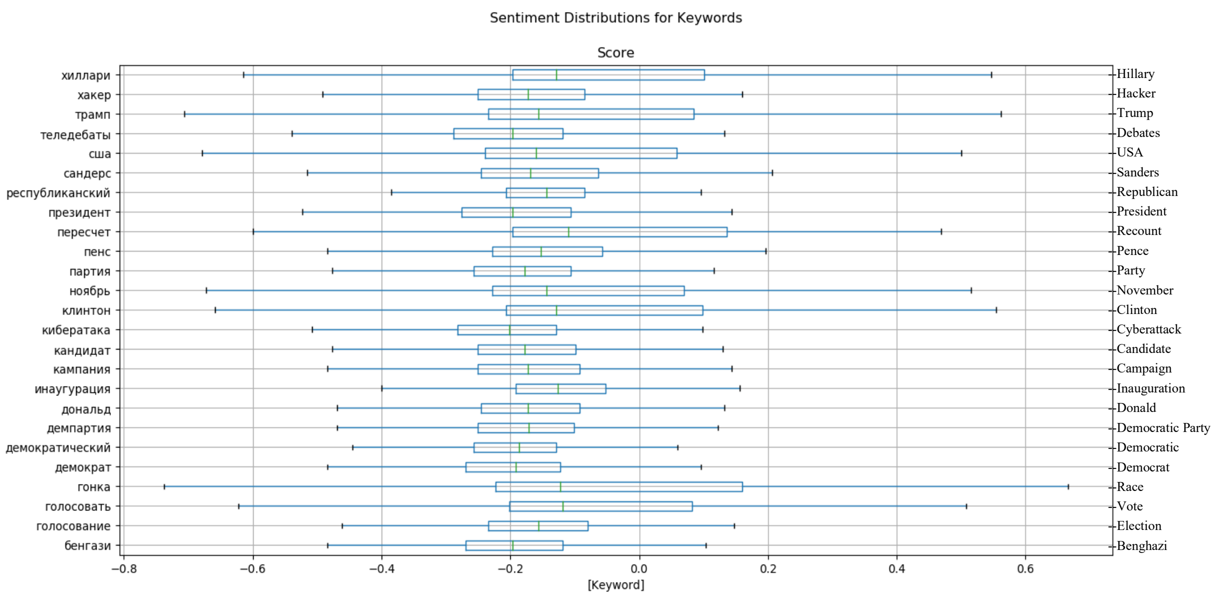

Figure 4. Sentiment Distribution for Each Keyword

Additional insight can be gained from visualizing the sentiment for individual words. As seen in *Figure 4,* only about 1/3 of the keywords have an interquartile range (IQR) that includes a positive polarity score. These keywords include *Hillary, Trump, USA, Recount, Clinton, November, Race, and Vote* (vb). Out of these keywords, *Race, Clinton, and Trump* have whiskers extending furthest into the positive, as well as whiskers past -.6. The keyword *race* could have a wide range of sentiment values due to the variety of contexts in which the word race is found. It is not necessarily unique to elections or politics.

A closer observation of both *Clinton* and *Trump* shows that each contain a wide range of sentiment and a virtually identical distribution. For *Clinton* the median was -.129, the mean was -.077, and the standard deviation was .182. For *Trump* the median was -.156, the mean was -.103, and the standard deviation was .193. This is interesting considering the expectation that one candidate would have more positive or negative coverage than the other. This expectation often happens because of how difficult it is for news media to escape bias or be completely objective, especially Russian media. As Erzikova (2014) explains, even regional outlets in Russia receive funding from the government and are thus pressured to align to a state agenda. And considering the geopolitical posturing between the US and Russia over their long history, one might expect Russia to overtly support one candidate over another.

This is not immediately visible just by analyzing at the distributions of sentiment values. Instead of observing a sharply contrasted and skewed distribution for one candidate, the distributions of these two keywords follow the corpus's pattern of a bimodal distribution with a skew towards the negative, with very few values of near-zero sentiment. At first glance, this distribution could suggest that Russian media does not cover them differently but treats them relatively equally regarding the tone and sentiment of their immediate textual context. The following sentences were selected from each polarity range based on observations of how they pertained to the topic at hand. They contain examples of both positive and negative media coverage of Trump and Clinton. Both candidates were susceptible to scandals, and both are often portrayed as competent and competitive candidates.

Random Sample of Negative Sentence Examples

связывать с электронный почта скандал вокруг клинтон разгораться, когда госдепартамент подтверждать, что она во время пребывание на пост госсекретарь пользоваться для решение государственный дело личный электронный почта и собственный почтовый сервер вместо официальный защищать правительственный электронный адрес.
The e-mail scandal around Clinton flared up when the Department of State confirmed that during her tenure as Secretary of State, she used her personal e-mail and her own mail server, instead of her government issued e-mail, to conduct state business.

кроме то, лидер президентский гонка среди кандидат-республиканец миллиардер дональд трамп открыто заявлять, что хотеть бы вернуть практика пытка вода или "что-нибудь плохо", сообщать rt.'
Additionally, the leader of the presidential race and Republican nominee, billionaire Donald Trump openly declared that he would like to return the practice of waterboarding or "something bad", reports RT.

Random Sample of Positive Sentiment Examples:

тогда и кремль и белый подтверждать, что владимир путин и дональд трамп договариваться о регулярный общение и в тот число о начинать подготовка в личный встреча.
Both the Kremlin and the White House confirmed that Vladimir Putin and Donald Trump would agree on regular communication and in that same vein will prepare for a personal meeting.

о душевный качество уважаемый миссис клинтон говорить ее послужной список.
Mrs. Clinton's track record speaks to her soulful qualities.

These randomly selected sentences show a few interesting things. First, it shows in the first two examples that the sentiment analyzer is sensitive to words like "scandal", "flare", and "bad", marking those sentences as negative. Second, it shows in the positively classified examples that some objectively neutral statements of fact or procedure are still classified with sentiment values. And third, it shows that perhaps some of the positive sentiments regarding *Clinton* are in sentences where others are telling about her positive qualities, or that what she is going to say will be positive in one way or another. This does not necessarily reflect the writer's perspective, or the agenda of the Russian media.

Despite the relatively equal distribution of sentiment between these two candidates, the difference in frequency of occurrence could imply an alternate agenda. Even though the ratio of positive to negative coverage for keywords was relatively equal, the sheer quantity of sentences could provide additional insight. For example, the number of times Trump appears in the corpus is roughly 3.5 times more frequent than how often Clinton is mentioned, whether positive or negative. The keyword *Clinton* is the 9th most common collocate of the word *Trump*, but *Trump* is the 2nd most common collocate of the word *Clinton*. This means that *Clinton* is rarely used in the corpus without referring to *Trump*, whereas *Trump* is used more frequently without use of the word *Clinton*. And while frequency is not necessarily proof of positive or negative bias, more news coverage is undoubtedly tied to a more strongly focused political agenda towards that topic. Referring to the research done by Field (2018), the number of mentions of the US in the corpus was inversely related to the economic trends in Russia when measured diachronically. His research cross-referenced the frequency data with a diachronic study of the Russia GDP. The research in this study is based completely on the content of the corpus and any comparisons with political or economic trends across time are beyond the scope of this study. Additional insight can be gained from looking at the sentences at the extremes of the sentiment polarity spectrum. Upon observation of the strength scores for the top 100 sentences that were positively classified for the keywords *Hillary, Clinton, Donald,* and *Trump*, it was apparent that there was a bias towards *Donald* and *Trump.* The strength scores for the 100 *Trump* sentences ranged from .53 to .72, with a mean of .58, a median of .56, and a standard deviation of .043. Meanwhile, the strength scores for the top

100 positive *Clinton* sentences ranged from .43 to .70, with a mean of .49, a median of .48, and a standard deviation of .057. A Kruskal-Wallis H-test was performed on these positive scores. An H-statistic of 97.11 and a p-value < .05 was found. The null hypothesis was rejected, finding a significant difference between the scores.

On the other hand, the negative sentence scores seem to slightly favor *Clinton* over *Trump*. *Clinton* had a range of -.52 to -.75 with a mean of -.56, a median of -.54, and a standard deviation of .04. *Trump* had a range of -.61 to -.78 with a mean of -.65, a median of -.64 and a standard deviation of .035. The same Kruskal-Wallis H-test was performed on the negative scores. An H-statistic of 118.748 and a p-value of < .05 was found. The null hypothesis was rejected, finding a significant difference between these scores as well.

However, these quantitative statistics for the strength scores do not tell the whole story. A closer qualitative observation of these sentence examples shows some unique patterns. Most of these positive sentences with *Trump* as the keyword were filled with praise for him and direct quotes of personal positive interactions with him. But for *Hillary* and *Clinton,* even though the sentences were classified with a strong overall positive-sentiment score, they were filled with negative sentiment directly for her. For example, the following sentence is positively classified, but not in *Hillary*'s favor.

однако сложный трамп в раз хорошо, чем радикальный хиллари
However, a complicated Trump is better than a radical Hillary

This sentence contains the keyword *Trump* as well, so it was also positively classified for the keyword *Trump.* There are other examples that expressed a strong positive sentiment, but a closer observation shows they specifically reflected negatively on the keyword in question. For example:

пустячок, а приятно, что все-таки не хиллари клинтон, который мы хорошо помнить в качество госсекретарь, с имя который связанный хаос и кровь в северный африка.
It's a small thing, but it's still nice that it's not Hillary Clinton, who we remember well as secretary of state, whose name is associated with chaos and blood in North Africa.

Similar patterns emerge when looking at the top 100 negatively classified sentences for these keywords. Extreme negative sentences containing the word *Trump* often focus on how Trump hates something or feels attacked by the media. Although the algorithm classifies this as negative, the observed resulting sentiment that is evoked in connection to the keyword is positive. For example:

трамп плевать на европа, а то более - на украина.
Trump 'spits on' Europe, and what's more – on Ukraine too

Here is a negatively classified sentence with both keywords *Trump* and *Clinton* (in contrast to the positive example containing both):

нет сомнение, что клинтон попытаться на дебаты представлять трамп женоненавистник.
There is no doubt that Clinton will try during the debates to portray Trump as a misogynist.

This example also shows how the sentiment analyzer picks up on the negativity surrounding the words "misogynist" but fails to account for the fact that the media is only reporting on how Clinton tries to portray Trump.

Interestingly, the negative sentences containing *Clinton* are almost completely uniform in their condemnation of *Clinton*. This is a shortcoming of the sentiment analyzer. It is unable to distinguish what parts of the sentence, which actors, and what direct objects are being positively portrayed. Instead, it classifies the text as one whole, unable to distinguish the more nuanced and specific sentiments a human classifier usually knows instinctually. The following sentences were each classified with negative polarity, but each contain more nuanced sentiments the automated sentiment analyzer was unable to identify.

многие считать, что клинтон - это совсем плохой, а трамп - плохой, но не очень.
Many think that Clinton is completely bad, and that Trump is just bad, and not very at that.

наследие клинтон трамп называть смерть, разруха, терроризм и слабость.
Trump calls Clinton's legacy one of death, devastation, terrorism, and weakness.

во время президентский дебаты в этот понедельник клинтон заявлять о недоверие к российский президент, и, как подчеркивать издание, этот недоверие взаимно.
During the presidential debate this Monday, Clinton declared her distrust of the Russian president, and, as the newspaper emphasizes, this distrust is mutual.

The last example is particularly interesting, as this is a rare instance of the media expressing an explicit opinion about one of the candidates. Plenty of examples in the corpus are comparable to the second clause of the sentence. In this clause a simple fact reporting takes place. The media can be selective about *which* facts to report, but the facts remain. It is in the final clause of the sentence where the opinion of the author comes into play. The author slips in an aside about a likely facet of the relationship between Clinton and President Putin. While reporting of quotes and events is quite routine in the corpus, it is rarer to encounter an overt and subjective opinion about politics.

To conclude the section on sentiment analysis, although the distribution of sentiment values between the chosen keywords (as well as throughout the entire corpus) does not vary significantly, the frequency with which the keywords appear in the corpus is significant. A qualitative analysis of the most extreme sentiment-filled sentences shows a pattern of negative bias towards *Clinton* and a positive bias towards *Trump*, often in direct contrast to the results of an automated sentiment polarity classification of the sentence.

3.3 – Word Embeddings

Another quantitative method of analyzing the corpus uses word embeddings/vectors. Word vectors use algorithms to capture a contextual and semantic value for words in the corpus (Ma & Zhang 2015; Bojanowski et al. 2017; Grave et al. 2018). Similar words, like colors of the rainbow for example, will have more similar vectors/embeddings when analyzed for context and frequency. In this study, the corpus was run through a python script using the Gensim Word2Vec (Rehurek & Sojka 2011) model to assign 100-dimensional vectors to each token in the corpus. The resulting vectors were then flattened into a 2-dimensional representation for ease of visualizing proximity.

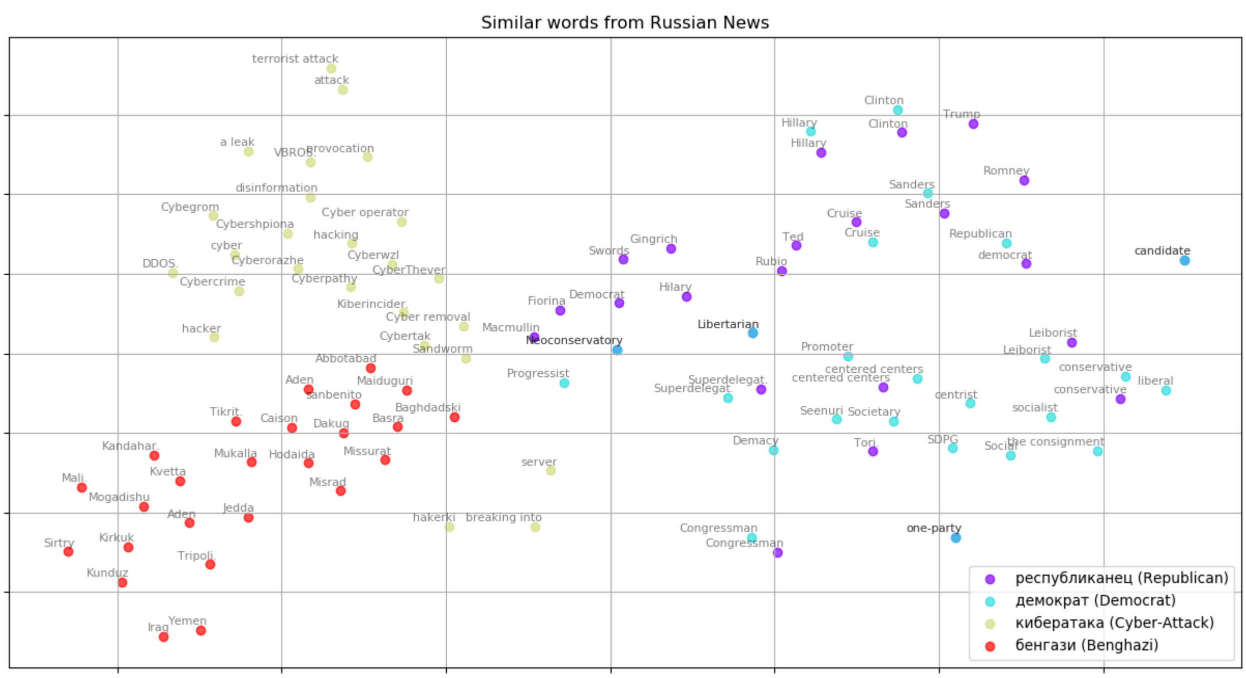

Figure 5. 2D Word Vector Visualization

Figure 5 shows comparisons between keywords (shown in the legend) and 25 words in the corpus that have the most similar vectors to the chosen keywords. In this example the chosen keywords were *Republican* (Purple), *Democrat* (Blue), *Cyber-attack* (Yellow), and *Benghazi* (Red). These keywords were chosen due to the expected polarity between the political parties, as well as their connection to separate scandals that happened or were brought up during the election. The vectors separated into roughly four quadrants, with a little more overlap of *Republican* and *Democrat* due to their shared political vector space. The horizontal and vertical axes are not specific to a category or value due to the complicated vector flattening from 100 dimensions to 2 dimensions. Nonetheless, the Euclidian distance between exemplars shows the relative similarity between words based on their contextual properties and usage in the corpus. The keywords *Benghazi* and *cyberattack* are very close in vector space due to both being events rather than political parties. And looking at the relation between the events and the political parties, the vector plot would seem to imply that the word *Cyberattack* mirrors more of the same vector/semantic properties as *Republican* than it does for *Democrat*, and vice versa for *Benghazi.*

However, *Cyberattack* does have three outliers that do cross into the bottom half closer to the *Democrat* quadrant. This could imply that while Benghazi was exclusively a Clinton scandal, the Cyberattack was painted more broadly by Russian media as an American issue. Indeed, many example sentences tend to downplay the cyberattack, prove Russia's innocence and lack of involvement, and label investigators of the cyberattack as unintelligent. Below are a couple of example sentences about the Cyber-attack.

власть сша неоднократно заявлять, что за различный кибератака стоять российский хакер, однако ни раз не предъявлять доказательство причастность россия к этот атака.

US authorities have repeatedly stated that a Russian hacker is behind the various cyberattacks, but have never presented evidence of Russia's involvement in this attack.

министр иностранный дело рф называть бредни сообщение западный, в тот число германский, сми о попытка россия дестабилизировать ситуация в фрг путем кибератака.

Russia's Minister of Foreign Affairs called the Western media's report, including the German one, about Russia's attempt to destabilize the situation in Germany using a cyberattack nonsensical.

An interesting feature of these sentences about the cyberattack goes back to sentiment analysis. Both sentences, although they are obviously downplaying the cyberattack and painting it in a negative light, are categorized with positive polarity scores by Dostoevsky (https://github.com/bureaucratic-labs/dostoevsky). In fact, they both appear in the top 100 positive sentences. However, upon observation of the sentiment strength scores, these sentences fall below .2 on the strength spectrum. Additionally, when the frequency in the corpus and sentiment histograms are analyzed, the frequency is far less than most of the other keywords on similar topics, and when it is talked about, the tone is more muted and neutral. It is apparent that Russia does not want to draw attention to the topic.

Another procedure for analyzing text with vectors involves vector math. As Allen & Hospedales (2019) explain in their research on understanding embeddings, vector math helps us utilize paraphrases in texts to answer analogical questions (Allen & Hospedales 2019). The classic example is "As man is to king, woman is to…" (Levy & Goldberg 2014), with the answer being "queen." Represented as a mathematical function, it looks like this:

$$\text{King - Man + Woman} \approx \text{Queen}$$

In other words, subtracting the semantic properties of *Man* that are encoded in the model from the word *King* and adding the properties of *Woman* yields a result of *Queen.* However, despite how elegant this method appears, it is not perfect and as Allen & Hospedales are clear to point out that while "embeddings factorize pointwise mutual information (PMI), it is *paraphrasing* that determines when a linear combination of embeddings equates to that of another word" (Allen &

Hospedales 2019: 1). So, to use this with the Russian news corpus, proper paraphrases must be discovered and analyzed. Additionally, the algorithm needed a few test cases to make certain it would output relatively accurate results from our corpus, as vector math often requires an extremely large corpus to function properly.

For the first test case, the paraphrase used was: "As *Clinton* is to *Democrat*, *Trump* is to…". The equation for the word embeddings looks like this:

$$Democrat - Clinton + Trump \approx$$

The Word2Vec model prints out the top most-similar words for the resulting vector as well as a similarity score (scale from 0 to 1 with 1 being completely equal and 0 being completely different vectors). The top result was *Republican* with a vector similarity score of .77. The second result was *Conservative* with a score of .7. This was to be expected as Clinton was the Democratic candidate and Trump was the Republican candidate, making for a simple logical paraphrase. The next test case used the paraphrase "As Moscow is to City, USA is to….". The top two results were *Country* (.53) and *Iraq (.51)*. Again, this was to be expected, as the USA is a specific of country just as Moscow is a specific city. Based on these two findings, it seems that the algorithm performs exceptionally when the topic is narrower and has a smaller vector space (i.e., politics). But despite being less precise for the Moscow/USA example, it is still accurate enough even with bigger ideas (like cities, countries, nations, etc.…) to be used for the purpose of this paper.

It would be desirable to compare 3 of the keywords together, like "Hillary is to Benghazi, as Trump is to…". But in the case of this study's chosen keywords, most of the keywords are nouns. It seems that subtracting nouns from different vector spaces, and nouns with complex, foreign, or multiple meanings do not perform well in this analysis. Adjectives or comparative nouns (like *Man* and *King*) work well in vector subtraction but attempting to subtract a vector like

Hillary from a complex and spatially foreign vector like *Benghazi* only seems to yield results for cities in Africa and the Middle East. *Benghazi* is especially difficult because not only is it a city in this corpus, but it is also the name of a political scandal utilized by multiple groups as political talking points. So, for this analysis, 2 keywords were used in the vector equations. *Clinton* and *Trump* were again chosen because of our interest in the media's sentiment towards the candidates. Additionally, the quality and consistency of the model in performing vector calculations yielded better results with these two keywords due to their symbiotic and opposing vector polarities.

A few of the most interesting vector calculations were based on the following paraphrases. The ideas behind these paraphrases came from a random sampling of sentences with these keywords.

Paraphrase	Result
"Clinton is to radical as Trump is to…."	*Romantic (.570), Pro-Russian (.568)*
"Trump is to strong, as Clinton is to…"	*Weak (.571), Sensitive (.527)*
"Trump is to victory, as Clinton is to…"	*Loss (.594), Defeat (.543)*
"Clinton is to Victory, as Trump is to…"	*championship(0.55), Draw(0.534)*

The first three examples show paraphrases of ideas common in the corpus. The fourth example is an attempt to perform vector math on an idea not found in the corpus. This attempt did not produce logical results. By comparing *Clinton* to *Victory,* an idea not found in the corpus, the algorithm did not know how to produce the analogous word to *Trump.* Instead of producing *Defeat,* the vector math produced words similar to *victory*. These vector paraphrases and calculations further support the evidence from the qualitative sentiment analysis that the Russian news media is biased in favor of Donald Trump. In many sentence examples, in collocates, and in word embeddings the Russian news media has wide coverage of both candidates, but that coverage is heavily skewed towards Trump and consistently uses language to vilify and undermine Hillary Clinton.

This chapter focused on sentiment, collocates, and word embeddings at more of a macro level due to the emphasis on quantitative analysis. The next chapter in this study will dig deeper into the corpus at a micro level, using a qualitative framework for discourse analysis to understand the how and why of this bias.

Chapter 4: Appraisal Framework and the Language of Geopolitical Agenda

In this chapter I provide a qualitative analysis of sample sentences from the corpus using an appraisal theory framework to answer the following research questions:

R2) How did the Russian media frame the candidates to Russian citizens?

R3) How do these trends reflect on possible Russian geopolitical agendas?

4.1 - Attitude

The first semantic category[1] in Appraisal Theory (Martin & White 2005), the category of Attitude, includes three sub-categories: Affect, Judgment, and Appreciation. Affect deals with the positive or negative feelings or emotions expressed or evoked by a text, such as what feelings are the Russian news media trying to evoke from their readers about Hillary Clinton. Judgment deals with the ethics of ideas or ideologies expressed by a text or appraiser, such as how competent of a leader the Russians think Donald Trump is. And Appreciation deals with the intrinsic or extrinsic value of an object of appraisal (Oteiza, 2017 p. 460), such as how much value the Russian news media attributes or appraises US candidates in comparison to their own beloved leader Vladimir Putin.

The sentences in this section were chosen because of how they represent the general trends/patterns discovered in the corpus through analysis of keywords in context as they pertain to possible Russian political agendas. More specifically these sentences show how the Russian

[1] The second area is Graduation, which is practically identical to polarity measures in sentiment analyses. Khoo (2012) explains that few researchers show interest in Graduation as it mainly pertains to strength of sentiment and is less befitting of a more qualitative analysis. Instead, the intensity of Graduation is often computed by counting the number of biased sentences in a text. This method would require a manual evaluation of each sentence in a sample or in the entire corpus depending on the scope of study. The results would be interesting to compare to the quantitative sentiment analysis performed in the previous chapter but is beyond the scope of this research. Regardless of the preferred method of analyzing Graduation, Chapter 3 is a satisfactory analysis of Graduation/polarity of sentiment in the corpus.

media seeks to evoke emotions, convey specific ideas about ethics and political ideologies, and make the reader consider the value of a particular topic or candidate.

не вынуждать констатировать, что в свой стремление к мировой господство для сша нет предел, они ничто не мочь останавливать.

It goes without saying that there is no limit to the United States' desire for world domination; nothing can stop them.

This example sensationalizes the idea of the US as a gluttonous superpower. It exudes an attitude of fear of the US to the Russian people and exaggerates the inevitability of US global domination. This type of rhetoric helps set the stage for more specific incidents played out in Russian news media, specifically regarding the election and US leadership. In the following examples, the Russian media portrays Trump as something to be feared. They attempt to evoke both feelings of fear and judgment in their appraisal of his competence.

более то, феномен трамп порождать сомнение в тот, в какой степень америка мочь и далеко осуществлять этот лидерство.

Moreover, the Trump phenomenon gives rise to doubts as to how far America can further exercise this leadership.

человек должный понимать весь серьезность использование этот оружие, у трамп нет абсолютно никакой понимание.

A person should understand the seriousness of using this weapon, but Trump has absolutely no understanding.

слово и действие г-н трамп демонстрировать неспособность терпеть взгляд, отличный от его собственный, вследствие что он впадать в ярость.

Mr. Trump's words and actions demonstrate an inability to tolerate attention given to anyone but himself, which causes him to fall into a rage.

This first example builds on the prior example about US world domination in seeding doubts as to the "Trump Phenomenon." Though the US has been known to be a global superpower, the media is perhaps implying that Trump is not up to the task to take the mantle of global leadership upon himself. The second example emphasizes that Trump "should understand" the ramifications of being in control of a particular weapon, but in reality he has "no understanding." This type of portrayal of an incompetent leader with access to nuclear weapons would undoubtedly sow fear among readers and would make them question whether it is ethical to deny great power to such leaders. The final example utterly demeans Trump's behavior to that of a child falling into tantrums and fits of rage. This type of sentence is a good example of a judgment and evaluative appraisal of Trump. To the Russian media and people, these qualities would seem out of place for a powerful leader, such as Vladimir Putin. An analysis of how Putin is appraised will be shown later in this study.

As shown in Chapter 3, Clinton often took the brunt of the negative news coverage, falling victim to the judgment of the media. But the examples above show that the Russian news media was often no more kind to Trump as it was to Clinton. Trump was often susceptible to judgment, negative news coverage, and reports of scandals. Both he and Clinton were often portrayed as incompetent leaders. Their appraised "value" is consistently lowered by the Russian media. This political rhetoric contrasts with coverage about Russian President Vladimir Putin, which, apart from frequently reporting on his daily routine, is quite often propagandist in nature and works to ever increase the appraisal value of Putin:

владимир путин быть и оставаться основатель и моральный лидер "единый россия", а партия являться центристский сила широкий народный большинство, опора президентский курс, инструмент ответственный правительственный политика, а также ключевой стабилизировать механизм социально-экономический ситуация.

Vladimir Putin was and remains the founder and moral leader of "United Russia", which party is the centrist force of a broad popular majority, the support system of the presidential course, the tool for responsible government policy, as well as the key to stabilizing the mechanism of the socio-economic situation.

исследование, проводить в август этот год показывать, что 85% республиканец находить путин сильный лидер, в тот время как только 18% придерживаться такой мнение относительно американский президент барак обама.

A study conducted in August of this year shows that 85% of Republicans find Putin a strong leader, while only 18% hold that opinion about US President Barack Obama.

This second example demonstrates that the Russian media, perhaps reflecting a cultural and societal viewpoint, values strong leadership. It conveys an attitude of disdain towards the US's "weak" leadership in comparison with the strong leadership of Russia. And though this could be interpreted as disdain for the entire US political arena, sentences like this guide the reader into focusing more and more on the Democrat/Republican divide. This divide is often accentuated using simple comparative methods. These methods guide the reader to favor one candidate over another by concentrating on the relationship that candidate has with a public figure the reader already admires (or has strong opinions about). In this case, Putin, who the media tends to value very highly, fills this comparative role:

думать, что путин совсем нелегко быть с хиллари. во всякий случай, трудно, чем с трамп

I think that Putin struggles to get along with Hillary. In any case, it is more difficult than with Trump.

Though not commenting directly on the election or US policy, this example uses Appreciation to evaluate the candidates. It presupposes that the reader has a favorable view of Putin, and therefore would have an unfavorable attitude towards Hillary or the candidate which Putin struggles to get along with. Because Putin has value in the eyes of the Russian media, a strong relationship with Putin is also portrayed as something to value. While the text does not overstate the tenuous relationship that Trump might have with Putin, it is clearly noted that Trump's relationship with Putin is, and would be, much stronger and stable than Clinton's. Thus, Trump is evaluated or appreciated higher than Clinton, but still has lower value than Putin in the eyes of the media, who are attempting to influence the Russian readers.

In general, the appraisal attitude of the corpus attempts to diminish the value of the United States as a whole and increase negative sentiment towards the candidates. But it is apparent that Clinton still receives much of the negative coverage and the authors attempt to associate more negative emotions (affect) with Clinton than with Trump.

4.2 - Engagement

The other semantic area of Appraisal Theory used in this study is the area of Engagement. This area deals with the commitments of the speaker. It deals with the question of whether the text is non-dialogic or dialogic, and how it contracts or expands dialog through various methods. A non-dialogic text expresses only one viewpoint or perspective as fact, does not provide evidence or research, and is usually a quick judgment appraisal or statement of an opinion. By contrast, a dialogic text expresses multiple perspectives or ideas, invites discussion of a topic, and does not offer a definite conclusion, but allows the reader to form their own conclusions. Considering the Russian Media Corpus, this section of the appraisal framework will help decipher the intents and agendas of the Russian Media in presenting information and ideas to the Russian people about the 2016 US Election.

In Chapter 3 it was shown how Clinton received a great amount of negative news coverage. This could potentially mean that there was simply more negative things and events to report about her. It does not necessarily mean the media was biased against her. Frequency does not necessarily prove intent. More convincing evidence is found in a qualitative analysis of the level of engagement found in the corpus. The nature of how the text engages with the topic will help determine the bias of the Russian media.

трамп с самый начало расцениваться как самый неординарный политик в история соединенный штат.
Trump was from the very beginning regarded as the most unorthodox politician in the history of the United States.

The above example forces the reader to consider how Trump is unorthodox, and why others regard him as such. It does not simply state that "unorthodox" is bad. It allows the reader to enter the dialog and ask "why?" and "is unorthodox such a bad thing?". The next example is more monoglossic as it presents only the viewpoint of the media and no quotes or references to the thoughts of others. It is classified with negative sentiment, but the negativity is insulated by Trump's "defiance" and "difference" from how typical US politicians are portrayed. And despite its monoglossic characteristics, it makes the reader think more deeply about why they should value a candidate who opposes typical US ideologies and defies the "Western political establishment."

трамп вызывающе неполиткорректный, пора грубый, даже внешне отличаться от штампованный клон западный политический истеблишмент.
Trump is defiantly politically incorrect, at times rude, and even outwardly different from the stamped clones of the Western political establishment.

By contrast, monoglossic sentences about Clinton are relatively non-dialogic. These sentences stymie attempts at dialog or critical thinking and provide an easy avenue for negative appraisals and evaluations of Clinton and the Democratic party.

а хиллари - настоящий террористка.
But Hillary is a real terrorist.

This examples efficiently guides the reader to pass quick judgment on the topic at hand. There is no room for dialog or discussion, only immediate appraisal. Hillary is made out to be a terrorist. The reader must either agree or disagree with the author's assessment because no other viewpoint, evidence, or discussion is presented.

By contrast, here is a sentence that at first glance might seem non-dialogic, but actually promotes a healthy dialog:

если хиллари собираться орудовать рапира, - предсказывать газета, - то трамп воспользоваться бутылка с отбивать дно.
If Hillary were going to wield a rapier – the newspaper predicts -, then Trump would use a bottle with a shattered bottom.

In this example the newspaper imagines a fight between Trump and Hillary. Hillary chooses a sword (rapier), while Trump is left with using a shattered bottle. It is a simple sentence that offers much to consider. The newspaper could be implying that Hillary is more sophisticated and prepared for a real fight, while Trump's competency only extends to brute force. On the other hand, the paper could be implying that Hillary is forced to use excessive force or cheat to even come close to beating Trump, who feels he can win without hardly lifting a finger.

However, very few sentences provide that kind of depth regarding Clinton. Most sentences about Hilary Clinton that are heteroglossic do not usually promote a positive dialog about her.

ранее кандидат в президент сша от демократический партия хиллари клинтон в ход дебаты заявлять, что кибератака на политический организация и институт соединенный штат, а также аккаунт частный лицо "очевидно, исходить от путин."

Earlier, during the debate, Democratic presidential candidate Hillary Clinton declared that the cyberattack on the political organization and institution of the United States, as well as on individual accounts, "obviously came from Putin."

The previous sentence embodies how the Russian media portrays Clinton and Putin's relationship. Clinton does not think highly of Putin and accuses him of a cyberattack. This type of reporting allows the reader to consider their own feelings towards Putin. Whether or not the reader believes a cyberattack happened and whether they believe it originated from Putin is beside the point. If the reader has a positive outlook on Putin, then Clinton accusing him of a cyberattack will naturally prompt the reader to adopt a defensive stance against Clinton to protect their view of Putin.

и вы увидеть то, что демократ тоже отмечать и, понятный, высмеивать: мировоззрение республиканец - "ни шаг назад", нельзя далеко терпеть весь то, что "настоящий" американец называть развал америка, торжество стиль жизнь, олицетворять демократ.

You will see what a democrat celebrates and, understandably, what they ridicule: the Republican worldview is "not a step back", you cannot long tolerate what a "real" American calls the collapse of America, which is the triumph of the lifestyle that personifies a Democrat.

This example is difficult to classify as either heteroglossic or monoglossic. It presents multiple ideas and evaluates both political parties and seems to promote some sort of dialog about the differences between the parties. However, it is apparent from a closer analysis of the text that their main argument is “Real” Americans are Republicans and they oppose Democrats because their lifestyle will bring about the collapse of America. If it does promote dialog, it is a shallow dialog and only seeks to widen the divide between the two parties.

избрание президент сша дональд трамп давать основание смотреть в будущее с осторожный оптимизм

The election of US President Donald Trump gives reason to look into the future with cautious optimism

In this example, the text presents a more open-ended monogloss. There are no quotes or external viewpoints from anyone other than the author. However, the phrase “cautious optimism” allows the reader to consider why they might be optimistic about Trump as the new US President, but at the same time exhibits caution. It does not explicitly tell the reader to be optimistic, but that that is a viable option for them. It similarly implies that Trump will be a competent leader, a friend to Russia, and has the approval of the Russian media, and therefore the approval of President Putin as well.

Overall, based on these samples from the corpus, it seems that the Russian Media seeks to promote more positive dialog about Trump and a future with him as a Russian ally. By contrast, the engagement with Clinton seems to be less dialogic and seeks to pass a quick negative appraisal of her candidacy for president.

The three semantic areas of appraisal – Graduation (using evidence from the Sentiment Analysis), Attitude, and Engagement - all appear to agree on a similar Russian agenda found in this corpus. As Oteiza states: “The basic reason for advancing an opinion is to elicit a response of solidarity from the addressee” (2017: 457). Consequently, Russian news reports about the United States are crafted with a general narrative or agenda. This narrative seems to emerge from the idea that Russia cannot ignore the United States. The United States is a world superpower and is often seen as a direct rival or competitor politically and economically. The presidential election is an ideal opportunity to amplify sentiment towards the US and specific candidates. Russia does not want to appear to overtly support one candidate over another, because more subtle tactics often work best in influencing public opinion. Additionally, Russia would not want to support the loser of the election, and so it plays it safe by using the media to further its agenda. Based on the observations of this study using sentiment analysis, attitude, and engagement, it is apparent that Russia favored Trump over Clinton and had hope that Trump would help Russia attain its goals, whereas it feared Clinton would provide more obstacles to Russia's goals. This agenda was translated into a narrative of a "destructive, weak, liberal" Clinton pitted against an "unconventional, rude, friend-to-Russia" Trump. While neither candidate for the US election was presented as ideal or even remotely comparable to Putin, supporting Trump and ridiculing Clinton fit the narrative the Russian media was trying to proclaim to the Russian people.

Chapter 5: Conclusion and Further Research

5.1 - Further Research

This study utilized a very large corpus (350 million words) to perform both a quantitative and qualitative analysis. More insight could be gained through better organization of the corpus itself. If the corpus were tagged for part of speech and organized by source, for example, it could provide insight into whether different news outlets were more or less biased, or whether they had different agendas or ideologies altogether. Additionally, a corpus organized by time of publishing, could provide greater insight. Kazun & Kazun (2017) explain how the Russian media portrayed Trump and Clinton differently in various stages: early in the campaign cycle, later in the campaign cycle close to the time of the election, immediately following the election, and months after the election. A diachronically tagged corpus would be extremely useful for researchers to compare how sentiment, frequency, and appraisal of the candidates and other political topics changes over time.

Additionally, the development of a new Russian sentiment analysis tool, trained on news article texts instead of social media posts from RuSentiment would likely be more sensitive in terms of the strength of polarity scores, although the distributions would likely be very similar. Should such a tool be created, it would be interesting to analyze texts at the paragraph and article level as well to see if more of the nuance in sentiment surrounding complex topics could be captured at different levels of tokenization.

Also, the very nature of this election in comparison to other US elections also presents some identifiable features that would be valuable to future researchers. A general attitude of misogyny from the Russian news media may have presented confounding variables into the dataset, which may not have been present in other US elections. Additionally, the relative frequency of mentions of Hillary Clinton in the corpus compared to the frequency of mentions of Donald Trump might give credence to the idea that "no news is good news." Perhaps Hillary Clinton was mentioned less because there was less negative news to report. Merely analyzing what was contained in the corpus could miss the value of analyzing what was omitted or ignored.

5.2 – Final Observations and Conclusion

This study used a corpus-based approach to research how the Russian news media portrayed the 2016 US Presidential Election to the Russian people. It used a large corpus compiled from various Russian news sites, using articles from 2016 to 2017. In Chapter 3, using several quantitative methods, an analysis of sentiment towards the two primary candidates and a few other topics was performed. Through this quantitative analysis, it was found that keywords associated with the presidential candidate Hillary Clinton as well as the cyberattack appeared to have a stronger negative bias in the corpus than keywords associated with Donald Trump or the Republican party. It was also discovered that although many sentences containing the keyword *Trump* were classified as negative, the negativity was not necessarily focused on *Trump* but rather on the keyword *Clinton*. Similarly, many sentences that were classified positively for the keyword *Clinton* portrayed Clinton in a negative light, despite the sentiment model classifying the entire sentence with positive polarity.

Similarly, an analysis of collocates, word embeddings, and word vector math calculations confirmed these findings of a negative bias towards Clinton. There were many examples of negativity towards the US in general as well as towards *Trump*. However, when analyzed as a whole, the sentiment of the corpus was skewed in favor of *Trump*.

In Chapter 4 a framework for discourse analysis using Appraisal Theory was used to analyze many sample sentences from the corpus. It explored how the Russian Media used semantic techniques in the appraisal categories of Attitude, Engagement, and Graduation to further their geopolitical agenda. In many sentences, an attitude of fear towards the US was conveyed and provided a foundation for judgment and evaluation of the candidates. Sample sentences showed how dialogic and non-dialogic methods were used to attempt to expand or contract dialog about specific topics or candidates. Sentences about Clinton were often non-dialogic, or the dialog was contracted due to extremes of appraisal and judgment. On the other hand, sentences about Trump were more often dialogic and attempted to expand acceptance of his divergence from the typical "Western Establishment."

From this analysis it is apparent that although the Russian domestic news media paints the US as a direct economic and geopolitical world competitor, it is deeply interested in the outcomes of US elections. While often appearing like a neutral observer, many examples in the corpus show overt subjective support for Trump over Clinton. Additionally, the quantity and frequency of news coverage about Trump is many factors greater than that for Clinton, and much of the strongest sentiment towards Trump is positive in nature, while the strongest sentiment towards Clinton is negative. This leads us to assume that it matters a great deal to Russia who wins the elections in the US. To get their own citizens on board, the media presents Trump as a candidate who has a good relationship with Putin, who is more likely to remove economic sanctions, who defies many Western ideologies, and who diverges from the typical US politician. By contrast, Clinton is portrayed as an enemy to Russia who is destructive to the US and the world at large. In terms of a generic Russian narrative and agenda, this study found that Russia does not seem to care who wins the US election for America's sake, but the Russian media uses the election to diminish the value the US, demean the candidates in the eyes of the Russian people, contrast them with the strong leadership of Vladimir Putin, and confirm Russia's superiority on the global stage.

References

Ädel, Annelie. 2010. How to use corpus linguistics in the study of political discourse. In *The Routledge handbook of corpus linguistics*, 591–604. Routledge.

Allen, Carl & Timothy Hospedales. 2019. Analogies Explained: Towards Understanding Word Embeddings. In *International Conference on Machine Learning*, 223–231. PMLR. http://proceedings.mlr.press/v97/allen19a.html (20 June, 2021).

Badawy, Adam, Emilio Ferrara & Kristina Lerman. 2018. Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 258–265. Barcelona: IEEE. https://doi.org/10.1109/ASONAM.2018.8508646. https://ieeexplore.ieee.org/document/8508646/ (25 January, 2021).

Bird, Steven, Ewan Klein & Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Bird, Steven, Edward Loper & Ewan Klein. 2009. *Python Module NLTK* (Natural Language Processing with Python. O'Reilly Media Inc.).

Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics* 5. 135–146. https://doi.org/10.1162/tacl_a_00051.

Boyd, Ryan L, Alexander Spangher, Adam Fourney, Besmira Nushi, Gireeja Ranade, James Pennebaker & Eric Horvitz. 2018. Characterizing the Internet Research Agency's social media operations during the 2016 US presidential election using linguistic analyses. PsyArXiv.

Burrett, Tina. 2018. Russian state television coverage of the 2016 US presidential election. *Demokratizatsiya: The Journal of Post-Soviet Democratization*. , The George Washington Institute for European, Russian, and Eurasian Studies 26(3). 287–319.

Darczewska, Jolanta. 2014. *The anatomy of Russian information warfare. The Crimean operation, a case study*. Ośrodek Studiów Wschodnich im. Marka Karpia.

Field, Anjalie, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky & Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. *arXiv preprint arXiv:1808.09386*.

Grave, Edouard, Piotr Bojanowski, Prakhar Gupta, Armand Joulin & Tomas Mikolov. 2018. *Learning Word Vectors for 157 Languages*.

Kazun, Anastasia & Anton Kazun. 2017. A Friend Who Was Supposed to Lose: How Donald Trump Was Portrayed in the Russian Media? *Higher School of Economics Research Paper No. WP BRP* 51.

Levy, Omer & Yoav Goldberg. 2014. Linguistic regularities in sparse and explicit word representations. In *Proceedings of the eighteenth conference on computational natural language learning*, 171–180.

Ma, Long & Yanqing Zhang. 2015. Using Word2Vec to process big text data. In *2015 IEEE International Conference on Big Data (Big Data)*, 2895–2897. https://doi.org/10.1109/BigData.2015.7364114.

Martin, J. R. & P. R. R. White. 2005. *The language of evaluation: appraisal in English*. New York: Palgrave Macmillan.

Oates, Sarah. 2014. Russian state narrative in the digital age: Rewired propaganda in Russian television news framing of Malaysia airlines flight 17. In *American Political Science Association Annual Meeting*.

Oteíza, Teresa. 2017. The appraisal framework and discourse analysis. *The Routledge handbook of systemic functional linguistics*. Routledge London 457–472.

Padda, Karmvir. 2020. Fake News on Twitter in 2016 US Presidential Election: A Quantitative Approach. *The Journal of Intelligence, Conflict, and Warfare* 3(2). 26–26.

Rehurek, Radim & Petr Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3(2).

Rogers, Anna, Alexey Romanov, Anna Rumshisky, Svitlana Volkova, Mikhail Gronas & Alex Gribov. 2018. RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian. In Emily M. Bender, Leon Derczynski & Pierre Isabelle (eds.), *Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018*, 755–763. Association for Computational Linguistics. https://www.aclweb.org/anthology/C18-1064/.

Rosenvall, James & Monte Shelley. 1988. *WordCruncher* (WordCruncher). BYU Office of Digital Humanities.

Rozenas, Arturas & Denis Stukal. 2019. How autocrats manipulate economic news: Evidence from Russia’s state-controlled television. *The Journal of Politics* 81(3). 982–996.

Shapiro, Adam Hale, Moritz Sudhof & Daniel J. Wilson. 2020. Measuring news sentiment. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.07.053. https://www.sciencedirect.com/science/article/pii/S0304407620303535.

Stokes, Ethan Christopher. 2017. *A global election: analyses of Chinese, Russian, and Saudi Arabian news coverage of the 2016 US presidential election*. University of Alabama Libraries PhD Thesis.

Suslov, Mikhail D. 2014. “Crimea Is Ours!” Russian popular geopolitics in the new media age. *Eurasian geography and economics* 55(6). 588–609.

Szostek, Joanna & Stephen Hutchings. 2015. Dominant Narratives in Russian Political and Media Discourse during the Ukraine Crisis. In, 183–196.

White, Stephen, Sarah Oates & Ian McAllister. 2005. Media effects and Russian elections, 1999-2000. *British Journal of political science* 191–208.