Lexical Choice in Presidential Debates

Jared Kelly

Abstract

This study addresses differences in the language use of Republican and Democratic Presidential candidates in election cycle debate transcripts. The primary research question in this analysis is whether Republicans and Democrats differ in terms of lexical choice, much in the same way that any two people from distinct backgrounds or different genders might, as Bamman, et.al. (2014) suggest in their analysis of twitter data. A unique corpus, The Presidential Debate Corpus (PDC), was created in order to explore these differences.

The PDC is composed from the transcripts of every Presidential (and Vice Presidential) debate since 1960. The corpus was annotated, then analyzed using AntConc, and its quantitative measure of "keyness" (determined by comparing frequencies with comparison corpora, using the statistical measures of Log Likelihood, Mutual Information and T-score). The initial analysis confirms a statistically significant difference in the lexical choices of Democrats and Republicans (e.g., Democrats use "country" while Republicans use "America" to refer to the same thing), and their speech conventions (e.g., in address, Democrats use more proper names, Republicans use more titles).

During this initial analysis, features were discovered warranting continued analysis of the corpus. Trends were noticed in the length of response and lexical choices used from the earlier to later decades. To analyze these trends, the data will be re-annotated for date. A longitudinal analysis was performed to identify keywords from each election cycle.

1 Introduction

At a time when politics have become so polarizing, I found myself interested in the differing mindsets of the left and the right. My first hypothesis was that the wants and desires and motivations of the two groups are not unlike each other and that a textual analysis of the two sides might show that they concentrate on the same areas of importance.

In order to test this hypothesis, I sought out a domain that would include balanced speech from both parties. The most instinctual was the domain of presidential debates. In a debate, the participants are led by a moderator in a balanced manner, attempting to ensure that both sides stay on topic and get equal opportunities for speaking. Upon searching, I was able to locate a website[1] that had collected all of the presidential debates back to 1960[2], between Kennedy and Nixon. I collected each of these debates separately in order to further process. Processing and collection details can be found in the next section, Corpus Construction and Processing.

Once the corpus was completely stored locally, but before I performed the processing required to divide it by speaker or party, I was left with 42 separate files; 1 each for 32

---

[1] presidency.ucsb.edu/debates.php

[2] In each of the next three presidential election cycles following the 1960 election, presidential candidates for one of the two major parties (Lyndon B. Johnson, 1964; Richard Nixon, 1968 & 1972) refused to participate (consensus is that they were hesitant after a supposed clear loss by Nixon in the 1960 debates) and so debates were not held.

Presidential debates, and 1 each for 10 Vice Presidential debates. The first round of analysis was performed on these 42 files, prior to processing. This analysis illuminated several “over-time” trends (detailed in the section on Analysis) that are interesting enough to warrant an extra step of processing later on, that is, maintaining the dates on the data so that we may notice how attitudes or topics have evolved over the last 60 years.

After this initial analysis, I annotated each utterance of each debate for both speaker name, party affiliation and year. Keyword analyses were performed comparing speakers, parties, and election cycles.

Each different participant’s speech was saved in a variable labeled by their name. Also, all moderators were stored as a gorup, and all question-askers were stored as a group. However, both of these two groups were stored into a common variable, “MOD”. Meanwhile, each speaker, along with having their text saved under their name, also had their text saved separately under their party designation, either “DEM” or “REP.” Thus, every spoken word was saved as either “MOD”, “DEM”, or “REP”. After this, I saved the different variables to different text files for further processing using the provided suite of tools discussed in the section on Analysis tools. In the next section, I will detail the steps used to process the text into this final state.

2 Previous Work

Almost all previous analysis of presidential (and other) debates is concerned chiefly with discourse analysis (Clayman, 1992; Morris and Johnson, 2011; Prabhakaran et al., 2014; Brown and Sovacool, 2017). Other work in the presidential debate realm has been concerned with the ramifications of presidential debates (Freeley, 1961), audience composition (Kenski and Jamieson, 2010), and citizen response (Shah et al., 2016), though always in the context of a single debate or election cycle. However, no previous work was found that attempts to collect all of the debates into a single corpus for analysis, and thus, no keyword analysis has been performed on the corpus until now. In their work analyzing keywords and language use in movie scripts, (Ramakrishna et al., 2107) use text to analyze psycholinguistic normatives, and other speaker characteristics using the LIWC tool. This paper extends the intuition of keyword analysis to the domain of presidential debates and the analysis thereof

Other research into lexical choice has revealed its usefulness as an indicator of topicality (Weng and Menczer, 2015). (Hampton and Shalin, 2017) consider lexical choice in tweets around the times of disasters, controlled with normal tweet patterns, and discover “interpretable patterns of language behavior” that correlate with current events of the respective regions of the users. This work reinforces the intuition that the most salient characteristics of a speaker, even current events in their region, can influence, and be predicted by, lexical choices made.

Several experiments have sought to train bag-of-words-based classifiers for detecting the gender of an author without ever reaching an accuracy greater than between 80%-90% (Plank and Hovy, 2015; Argamon et al., 2007;).Consistent accuracies indicate that while it is possible to use lexical choice to predict gender of an author, there is some interaction of other factors which keep gender from being completely predictive on its own.

Bamman et al (2014), take experimentation a bit further by classifying text in an unsupervised manner first, and then trying to determine what patterns of demographics appear in each of the unsupervised clusters. This is a reversal of the trend to try to read gender onto text and instead revealed that there are far more factors related to a person’s speech than gender. In fact, they argue for an interaction of factors which include gender, age, race, and other personal characteristics.

While one might argue that someone's political stance is not as personal a characteristic as say race, gender, or age, others might argue that it is more consistent and tightly-held to some than their age (which is always increasing) or gender-identity (which also evolves). For this reason, especially in the domain of Presidential Debates (where participants are typically lifelong politicians holding to a certain set of beliefs), it stands to reason that someone's political stance could have a noticeable effect on lexical choice.

Antconc uses the log-likelihood of terms in a pair of documents in order to determine keyness, or the effect size of the difference in the use of a term between a target corpus and a source corpus. We use this measure to find words which occur far more often in one corpus than another, such as a Republican corpus and a Democrat corpus.

3 Corpus Construction and Processing

The debates were all scraped from the American Presidency Project[3]. A parsing script was developed to split each debate into text files of different configurations for use downstream as input to AntConc (see Experiment). The output of the parsing script is a series of .txt files, divided by speaker. There are two types of output files (containing every token spoken by a particular subset of speakers):

- 1 file per candidate (e.g Bush, Obama, etc.)
- 3 files per election cycle, containing the year in the name as well (REP, DEM, and Moderator)[4]

These txt files were then used as input to the Corpus Analysis tools discussed in the next section.

To perform analyses, I used several different corpus tools. The first tool used was Laurence Anthony's AntConc. This program allows wordlist creation and comparison, as well as exploration of concordances, N-grams, and collocates, all across multiple files. This made it especially useful for analyzing the PDC with its .txt format. All processing of the data was done in Pycharm, using the Python programming language. Part-of-speech tagging was done using Spacy.[5]

4 Results and Analysis

The analysis found several keyword terms for each subgroup, and a visual inspection of the results confirms the intuition of the approach, that we should be able to identify keywords of topicality from each sub-group. The most effective demonstration of this is observed at the election-cycle keyword level.

4.1 Analysis by Election Cycle

In order to find keywords of relevance, we used a standard English stop word list to filter out irrelevant function words. Once the results were processed, the strongest keywords from each debate were names of candidates. Candidate names were further filtered, but other names (e.g. Putin, Bin Laden, etc.) were left, since they were clearly topics, and not participants. The remaining top keywords for each election cycle are presented in Table 1.

---

[3] https://www.presidency.ucsb.edu/

[4] In 1992, there were 4 files, since there was an IND category as well, containing the speech of Ross Perot and his running mate, Admiral James Stockdale.

[5] https://spacy.io/

Table 1: Topic Keywords by Election Cycle – Stop words and candidate names filtered out.			
1960	**1976**	**1980**	**1984**
communists	Watergate	treaty	Nicaragua
islands	inflation	oil	soviet
Formosa	Vietnam	OPEC	Lebanon
1988	**1992**	**1996**	**2000**
conventional	admiral	capital	seniors
hostages	character	welfare	prescription
contra	trickle	scheme	surplus
2004	**2008**	**2012**	**2016**
terrorists	Afghanistan	businesses	ISIS
terror	energy	Obamacare	Russia
homeland	Pakistan	Benghazi	emails

As you can see from the table, the remaining keywords are clearly indicative of either policies or events occurring around the time of the debate, or during the previous presidency. A brief review of only the last four cycles reveals many items of memorability, including ISIS and Russia (2016), Obamacare and Benghazi (2012), Afghanistan and energy (2008), and finally terrorists, terror, and homeland in 2004. It appears as though for a time in the 1990's the topics were more concerned with our own country and our own people and problems, but in the decades before and since, there is much more discussion of foreign affairs and issues.

4.2 Analysis by Political Party

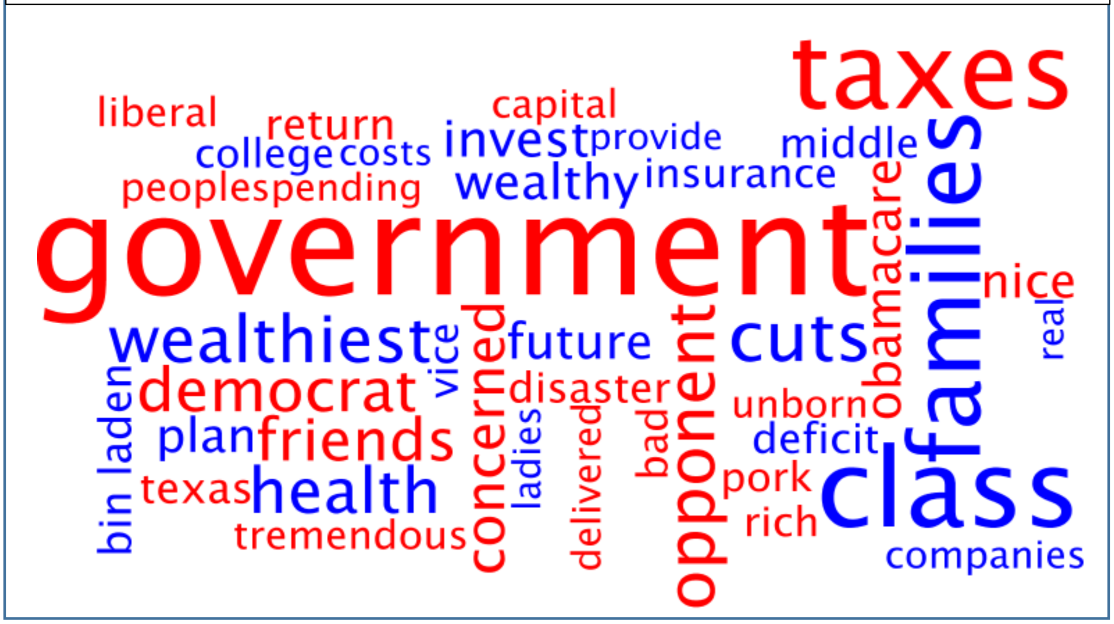
Next, analysis was performed to compare the overall language use of republicans with the overall language use of democrats. Again, stop words were filtered out to avoid unnecessary noise in topic keywords. The results of this keyword analysis are presented in a word-cloud in Figure 1.

My initial hypothesis was that the two groups would not differ greatly in their word use. My justification for this was that the debate domain and its structured nature would skew speech towards the same topics and words. Contrary to my hypothesis, there were indeed largely different keywords used between the two groups.

Keywords in red include things that we are unsurprised to see the Republicans bemoaning with greater frequency. For instance, "government", "Obamacare", "spending", "opponent", "unborn", and more. We even find "liberal" and "democrat" in the list here. Conversely, while the Democrats do also have some keywords that are probably things they dislike (e.g. "bin laden", "wealthy", and "class"), there are far more "positive" sentiments in their list, or topics that they are in favor of. This includes keywords such as "health", "college", "insurance", "future" and "families".

Of note is the observation that neither "Republican" nor "conservative" are found as democratic keywords. This is further indication that speakers on the left prefer to discuss positivity, and to report the virtues of good policies and ideas, whereas speakers on the right prefer to grumble over past policies or interpreted failures. One final observation is that when we expand the search to include keywords that fit $p < .05$ (instead of $p < .01$), we note that the

Figure 1: Republican – Democrat Keyword Comparison Results. Blue = Democratic Keywords. Red = Republican Keywords. Size ≈ Log-Likelihood effect size. For all terms, $p < .01$



keyword "country" appears in the Democratic list, and "America" appears in the Republican list, perhaps suggesting that either A) they use the two different terms to refer to the same entity (which also implies some sort of difference in view), or B) Democrats are more concerned with things external to the US, and Republicans are more concerned with things internal, and less worldwide.

Since this analysis was performed at the entire corpus level, a pertinent next step would be to perform a similar analysis at each election cycle level, comparing the two groups at 4 year intervals instead of overall.

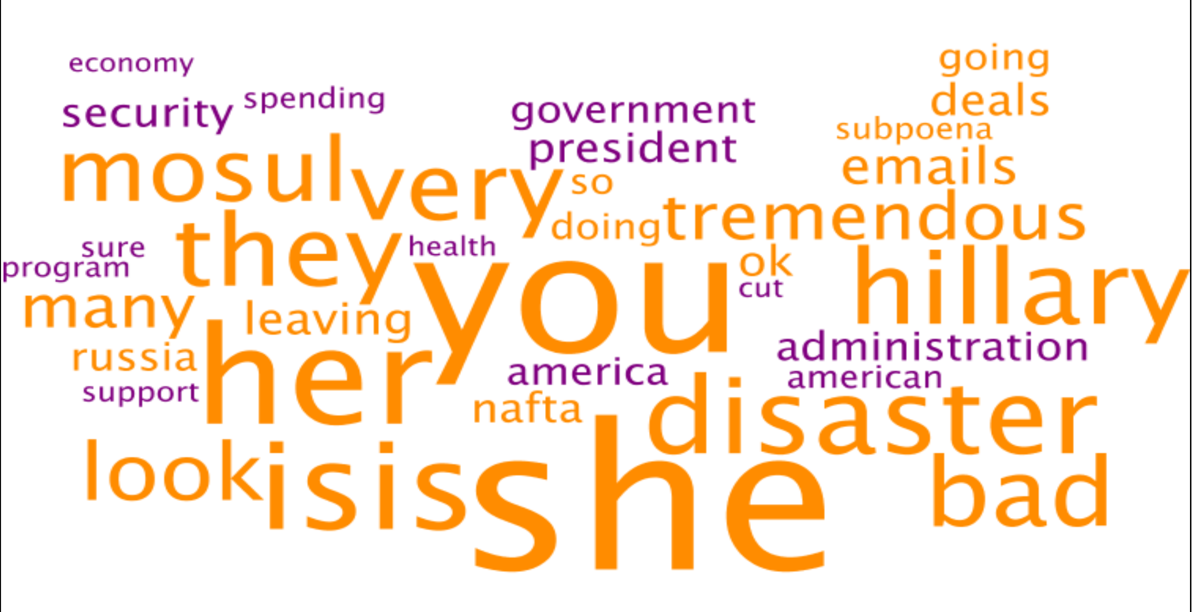
Before I removed stop words from the Republican-Democrat analysis, I noticed the presence of the female pronouns (she and her), which were also heavily prevalent in the 2016 election cycle keyword analysis (prior to stop word removal). It was at this point that the speech of a single particular outlier (Donald Trump) in the corpus was noticed to be skewing much of the rest of the Republican data. At this point, an analysis was performed comparing Donald Trump's language use to the rest of the candidates in the corpus.

4.3 Analysis of Donald Trump

A tri-partite analysis was performed on the speech of President Donald Trump, incorporating keyword analysis, a chi-squared analysis of his adjective use, and a Type-Token Ratio (TTR) analysis of all candidates in the corpus, highlighting the simplicity of his speech when compared with previous candidates.

Keywords The first step in the analysis of the speech of Donald Trump was a straightforward keyword analysis in the same manner as the rest, using the log-likelihood metric from within AntConc. The results of this initial analysis are found in the word cloud in Figure 2. The strongest keyword "she" in Donald Trump's speech is more than just a product of the fact that he is the first candidate to face a female presidential candidate, as well as the first to debate a female three times (Geraldine Ferraro and Sarah Palin both debated, but only once, as VP candidates). I

Figure 2: Trump vs all others Keyword Cloud. Orange = Trump key terms, Purple = other candidates, Size ≈ Log-Likelihood effect size ("she" had its effect size reduced by 75% so as to not skew the other terms in the figure so small they couldn't be read). For all terms, p < .01



contend that this is more of what was previously discussed before, the tendency of a Republican to stick to complaining about the negative, instead of concentrate on the positive. It is clear from these keywords that he spends far more time talking about his opponent, than about anything he is hoping to improve. The fact that "you" is the second strongest keyword, and "her" is the third, followed by "ISIS," "disaster," and "Hillary", further corroborates this idea.

Of particular interest is that "spending", "government", "American", and "American" all appear in Donald Trump's negative keywords, meaning that others use the terms far more than he. In this way, he follows Democratic trends. In fact, after processing the data without Donald Trump's speech, words like "bad" and "tremendous" disappear from Republican keyword lists. His use alone is enough to make them keywords for all Republicans.

Adjectives Because of growing interest in the speech patterns of the president, and specifically because of the previous results identifying Trump as an outlier, further analysis was performed on the corpus, specifically in the domain of adjective use.

Table 2: Total Adjectives and Tokens

Token Type	Trump	Other	Totals
Adjectives	1822	44673	46495
Remaining	29692	482631	512323
Total Tokens	31514	527304	558818

Table 3: Adjective Use by Trump – Subset of Adjectives, Freq/10k (raw freq.)

	Trump	All Others	Totals
bad	15.5 (49)	1.7 (92)	141
tremendous	8.9 (28)	0.7 (36)	64
great	17.1 (54)	7.9 (415)	469
inner	5.7 (18)	0.7 (35)	53
horrible	2.9 (9)	0.1 (7)	16
Total Tokens	31514	527304	558818

My original hypothesis was that Trump used far more adjectives than other candidates. After all, there are certainly adjectives such as “great”, “tremendous” and “wrong” that Donald Trump is quite notorious for using excessively.

To facilitate this analysis, the individual speaker corpora were annotated for part-of-speech and processed to count each token and its tag. Table 2 shows that contrary to original thought, Trump’s adjectives account for only 5.78% of total tokens, compared to 8.47% of total tokens in the rest of the corpus (x^2= 282.18, *df*=1).

Despite the fact that he uses significantly fewer overall adjectives than the other speakers, he still uses certain adjectives far more than any previous candidate. Table 3 shows five different adjectives that were pulled from Donald Trump’s keyword list for further analysis (all of them are significant to p < .001). The lowest x^2 value of any of the terms is “great” with a value of over 30, well past significant for the chi-squared distribution.

Since Donald Trump’s total tokens only amount to a small fraction when compared with the rest of the corpus, the adjective counts were normalized to frequency/10,000 tokens. Once normalized, the disparity in the use of these particular adjectives becomes clear. Donald trump uses the rather elementary adjectives “great” and “bad”, and 22X more than other candidates, respectively. It is even more alarming when one considers that these numbers are in comparison with the sum total of all other candidate speech, and not an average. He uses horrible almost 29X times more than all of the rest of the candidates (9 times to a combined 7 for the rest). The fact that he is using so few adjectives, and yet uses certain relatively elementary adjectives so incredibly much, indicates a trend to overuse simple terms. To confirm this, a final analysis was performed to calculate simple Lexical Diversity, using Type-Token Ratio.

Table 4: Presidential Debate Corpus Speaker Descriptive Statistics – Sorted first by number of debates, then by TTR. Red = Republican, Blue = Democrat, Black = Independent.

Year	Candidate	Types	Tokens	N	TTR
84, 88, 92	Bush (41)	3229	33273	6	9.7%
00, 04	Bush (43)	3431	42129	6	8.1%
08, 12	Obama	3501	45555	6	7.7%
92, 96	Clinton (W)	2786	30116	5	9.3%
92, 96, 00	Gore	3908	46916	5	8.3%
76, 80	Carter	2794	22414	4	12.5%
60	Kennedy	2026	18402	4	11.0%
60	Nixon	1936	18057	4	10.7%
04	Kerry	1392	7762	3	17.9%
76	Ford	1923	13898	3	13.8%
76, 84	Mondale	2224	16092	3	13.8%
92	Perot	2020	15396	3	13.1%
80, 84	Reagan	2327	17818	3	13.1%
16	Clinton (H)	2404	20044	3	12.0%
08	McCain	2411	20800	3	11.6%
76, 96	Dole	2283	21596	3	10.6%
12	Romney	2281	24266	3	9.4%
16	Trump	1990	23424	3	8.5%
00,04	Cheney	2014	13084	2	15.4%
88, 92	Quayle	1520	10861	2	14.0%
08, 12	Biden	1923	14559	2	13.2%
88	Dukakis	1785	14514	2	12.3%
92	Stockdale	672	2286	1	29.4%
84	Ferraro	1192	6044	1	19.7%
96	Kemp	1352	7037	1	19.2%
12	Ryan	1358	7456	1	18.2%
88	Bentsen	1105	6108	1	18.1%
16	Kaine	1417	8084	1	17.5%
16	Pence	1410	8090	1	17.4%
08	Palin	1352	7978	1	16.9%
04	Edwards	1279	8432	1	15.2%

Type-Token Ratio The final analysis performed was a simple type-token ratio breakdown for each candidate in the corpus. Because of the disparity in the opportunity that different speakers had to participate, Table 4 has been sorted first by number of debates participated in, then sorted by TTR. Of the candidates who participated in three different debates (and so had equal opportunity for speaking, making TTR relevant), Donald Trump has the lowest TTR score by almost a full point. His opponent in 2016, Hillary Clinton, was almost 30% more lexically diverse than him in terms of this same metric. The only speaker in the corpus with a lower lexical diversity score than Donald Trump (8.5%) was Al Gore (8.3%), and Gore participated in more total debates (including 2 cycles as a VP candidate), and used far more tokens, skewing the metric, which makes this an unfair comparison.

Overall, these results indicate that Donald Trump speaks in a significantly less diverse, and more simple manner, and that he concentrates more on speaking about others and other problems, than focusing on improvements or policies of his own. This even includes focusing more attention on Hillary Clinton than a wall.

5 Future Research

Next, I would like to process the data through the Spacey module in Python, in order to be able to further inspect it for counts of Named Entities, sentences, and the like. Also, I would like to experiment with different types of classifiers, such as bag of words, N-gram language models, Part of Speech models, and stop-word counting models. This to see how the differences between Republican and Democratic speech compare with other observations (i.e. Is there more or less disparity between Republican and Democratic speech that there is between negative and positive movie reviews on IMDB, or than between the English and Spanish languages?

Finally, I would be interested in performing a complete analysis similar to that of (Ramakrishna et al., 2107) in order to perform analyses using psycholinguistic normatives and the Linguistic Inquiry and Word Count (LIWC) tool to quantify speech patterns in different domains, such as emotion, or age of acquisition.

6 References

Laurence Anthony. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Shlomo Argamon, Moshe Koppel, James W. Pennebaker and Jonathan Schler. 2007. Mining the Blogosphere: Age, gender and the varieties of self-expression. *First Monday*, 12(9).

David Bamman, Jacob Eisenstein and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135-160.

George Brown and Benjamin K. Sovacool. 2017. The presidential politics of climate discourse: Energy frames, policy, and political tactics from the 2016 Primaries in the United States. *Energy Policy*, 111:127-136.

Jessica Autumn Brown. 2016. Running on Fear: Immigration, Race and Crime Framings in Contemporary GOP Presidential Debate Discourse. *Critical Criminology*, 24(3):315-331.

Steven E. Clayman. 1992. Caveat orator: Audience disaffiliation in the 1988 presidential debates. *Quarterly Journal of Speech*, 78(1):33-60.

Lisa A. Fast and David C. Funder. 2008. Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, 94(2):334-346.

Austin J. Freeley. 1961. The presidential debates and the speech profession. *Quarterly Journal of Speech,* 47(1):60-64.

Louis August Gottschalk and Goldine C Gleser. 1969. *Measurement of psychological states through the content analysis of verbal behavior*. California Univ. Press, Berkeley, edition.

Andrew J. Hampton and Valerie L. Shalin. 2017. Sentinels of Breach: Lexical Choice as a Measure of Urgency in Social Media. Human Factors: *The Journal of the Human Factors and Ergonomics Society*, 59(4):505-519.

Kate Kenski and Kathleen Hall Jamieson. 2010. Presidential and Vice Presidential Debates in 2008: A Profile of Audience Composition. *American Behavioral Scientist*, 55(3):307-324.

Eric Morris and Jessica M. Johnson. 2011. Strategic Maneuvering in the 2008 Presidential Debates. *American Behavioral Scientist*, 55(3):284-306.

James W. Pennebaker, Matthias R. Mehl and Kate G. Niederhoffer. 2003. Psychological Aspects of Natural Language Use: Our Words, Our Selves. *Annual Review of Psychology*, 54(1):547-577.

Barbara Plank and Dirk Hovy. 2015. Personality Traits on Twitter. *In Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 92-98, Lisboa, Portugal.

Vinodkumar Prabhakaran, Ashima Arora and Owen Rambow. 2014. Staying on Topic: An Indicator of Power in Political Debates. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP).

Dhavan V. Shah, Alex Hanna, Erik P. Bucy, David S. Lassen, Jack Van Thomme, Kristen Bialik, JungHwan Yang and Jon C. W. Pevehouse. 2016. Dual Screening During Presidential Debates. *American Behavioral Scientist*, 60(14):1816-1843.

Lilian Weng and Filippo Menczer. 2015. Topicality and Impact in Social Media: *Diverse Messages, Focused Messengers*. *PLOS ONE*, 10(2):e0118410.