

|  |  |
| --- | --- |
| Morphosyntactic Analysis of Democratic and Republican Political SpeechesMSBA Capstone Project | AbstractWeI analyze five sessions’ convention speeches of the Democratic and Republican Parties from 2004 to 2020 and 1,038 presidential speeches from 45 U.S. Presidents from 1789 to 2021, from George Washington to Joe Biden. Two research approaches are applied to the textual analysis: topic modeling and permutation testing. We find that over the years the Democratic Party has had more topics related to energy, economy, healthcare, and the future while the Republican Party has had more topics related to nation, business, and children. From our permutation tests, we find that the Republican Party is more likely to use past tense in both corpora and more likely to use singular pronouns in convention speeches. eI analyze five sessions’ convention speeches of the Democratic and Republican Parties from 2004 to 2020 and 1,038 presidential speeches from 45 U.S. Presidents from 1789 to 2021, from George Washington to Joe Biden. Two research approaches are applied to the textual analysis: topic modeling and permutation testing. We find that over the years the Democratic Party has had more topics related to energy, economy, healthcare, and the future while the Republican Party has had more topics related to nation, business, and children. From our permutation tests, we find that the Republican Party is more likely to use past tense in both corpora and more likely to use singular pronouns in convention speeches. WeI analyze five sessions’ convention speeches of the Democratic and Republican Parties from 2004 to 2020 and 1,038 presidential speeches from 45 U.S. Presidents from 1789 to 2021, from George Washington to Joe Biden. Two research approaches are applied to the textual analysis: topic modeling and permutation testing. We find that over the years the Democratic Party has had more topics related to energy, economy, healthcare, and the future while the Republican Party has had more topics related to nation, business, and children. From our permutation tests, we find that the Republican Party is more likely to use past tense in both corpora and more likely to use singular pronouns in convention speeches. Bu, XinMSBA Candidate College of Business University of Montana |

Spring 2023

Contents

[Introduction 1](#_Toc130753559)

[Review of Related Literature 3](#_Toc130753560)

[Methods 7](#_Toc130753561)

[The datasets 7](#_Toc130753562)

[The research approaches 9](#_Toc130753563)

[Topic modeling 9](#_Toc130753564)

[Permutation test 10](#_Toc130753565)

[Results and discussions 12](#_Toc130753566)

[Topic modeling 12](#_Toc130753567)

[Permutation test 14](#_Toc130753568)

[Conclusion 17](#_Toc130753569)

[References 18](#_Toc130753570)

[Appendix A 20](#_Toc130753571)

[Appendix B 21](#_Toc130753572)

[Appendix C 26](#_Toc130753573)

[Appendix D 30](#_Toc130753574)

[Appendix E 34](#_Toc130753575)

Introduction

Language in a political speech fulfills a specific linguistic function within its given context. Linguistic structures not only serve to deliver relevant messages to the audience, but also are embedded in the broader societal and political context. Linguistic analysis of political discourse is most successful when we relate the details of linguistic behavior to political behavior (Schaffner, 1996). The linguistic behavior this paper aims to examine is tense and first-person pronouns, in the context of Democratic and Republican convention speeches. Tense and first-person pronouns represent linguistic dimensions of time and space, two deictic functions (Chilton, 2004). Therefore, we will explore these linguistic micro-level structures to analyze political speech textual data, in addition to topics in each year of each party, to further understand the underlying societal and political implications.

According to Charteries-Black (2018), the two main types of political speeches are policy-making and consensus-building. A policy-making speech addresses aspects of policy in the near future, whether on war, immigration, gun control, or student debt. A consensus-building speech is made to inspire, motivate, and unite the audience to shape shared beliefs between the speaker and audience. In party convention speeches, the focus of our study, the distinction between the two types is not clear-cut. Convention speeches incorporate a speaker’s political views to build connections with the audience. They also intend to influence voters’ decisions in the presidential election. Held every four years, Democratic National Convention (DNC) and Republican National Convention (RNC) provide a platform for both parties to deliver their ideology to their audience, in order to persuade their opponents or undecided voters to be supporters, or to motivate their existing support base for continued support.

Analysis of political speeches traditionally applies rhetoric or discourse approaches (Charteries-Black, 2018). A recent trend using natural language processing (NLP) techniques to analyze political speech has emerged in the past two decades. Major NLP techniques have been used to explore the linguistic meaning of speeches. In our paper, in addition to using topic modeling (an NLP technique) to track the evolution of topics between 2004 and 2020, we propose a permutation test to investigate the morphological structure of convention speeches, specifically the use of verb tense and pronouns. We are interested in gaining insights about the differences in speech patterns between parties at the subtle linguistic granularity level.

The concept of verb tense deals with time. From a past-oriented vs. future-oriented point of view, tense contributes to interpreting Democratic and Republican Party ideology. The Democratic Party is well known for its progressive ideology compared to conservative Republicans, who are more past-oriented. For conservatives, the present is more likely built upon the past. Conversely, the past proves the present. In contrast, progressiveness tends to use the future to prove their present stance.

First-person pronouns imply power, social distance, and social imposition. When we use “I”, we likely show authority, personal responsibility, and distance ourselves from others. When we use “we”, we indicate closer social distance, invite audience involvement, and welcome shared responsibility. Similarly, as stated by Chilton (2004):

*“Pronouns are one class of words that can perform deictic functions. For example, in political discourse the first person plural (we, us, our) can be used to induce interpreters to conceptualize group identity, coalitions, parties, and the like, either as insiders or outsiders” (p.56).*

Moreover, the first pronoun singularity and plurality are associated with individualism and collectivism, respectively. Kashima and Kashima (2003) supported the claim that overall pronoun use is correlated with individualism. The use of the first pronoun singular “I” indicates independent or individual self while the plural “we” indicates collective or interdependent self (Uz, 2014).

Given the significance of studying tense and first person pronouns in Convention speeches, we propose the research questions as below:

* How does the first person pronoun usage vary by party?
* How does the future tense usage vary by party?

The paper is structured as follows. Section 1 provides a review of related literature on convention speeches, political speeches, and natural language processing, with a focus on methodology. Section 2 describes the dataset and methodological procedures. Section 3 presents and interprets the results. Section 4 gives recommendations for future research and concludes the paper.

# Review of Related Literature

This review synthesizes literature related to two topics: convention speeches for the Democratic and Republican parties over the past two decades, and NLP and its application to political speeches. The sources of our literature include peer-reviewed academic articles, published books, and book chapters. We used the following keywords to narrow down the literature search: “convention speeches”, “political”, “Natural Language Processing”, “part-of-speech”, and “parsing”. Overall, the literature consists of two parts: relevant studies using convention speeches or political speeches as their datasets, and theories related to political speeches and NLP.

We selected the relevant studies based on the following criteria: first, the speeches in the studies were verbally delivered to an audience. Second, the speeches were delivered by American politicians in a monological form, meaning that there were no direct interactions between two opponents. Third, the speeches carry political views.

A handful of studies have examined Democratic and Republican Party presidential nomination convention speeches (Alvi Baseer, 2011; Benoit, Blaney Pier, 2000; Deason Gonzales, 2012; Frank McPhail, 2005; Kendall, 2017; Selby, 2013; Shekels, 2017; Vianica Tanto, 2021; Vigil, 2014). These studies analyzed convention speeches through different lenses: feminine (Vigil,2014),  moral (Deason Gonzales, 2012), linguistic (Alvi Baseer, 2011; Benoit, Blaney Pier, 2000; Frank McPhail, 2005; Selby, 2013; Shekels, 2017; Vianica Tanto, 2021), and political views (Holbert, Hardy LaMarre, 2017; Kendall, 2017), but none have used computational methods from NLP to study convention speeches. From a feminine point of view, Vigil (2014) examined speeches delivered by the nominees’ spouses from 1992 to 2012 and found that the potential first ladies’ speeches were largely restricted to their home and family, as well as their supporting roles in shaping their husbands’ political future. Deason and Gonzales (2012) addressed how Democratic and Republican politicians varied in their adoption of the metaphor of a nation as a family. The concept of family has shaped the political ideologies of both parties, further facilitating persuasion in the 2008 presidential convention acceptance speeches.

Other studies that analyzed convention speeches chose to use an individual’s speech as a sample, such as Obama (Alvi Baseer, 2011, Frank McPhail, 2005), Biden (Vianica Tanto, 2021), and Huckabee (Selby, 2013), applying traditional rhetoric or discourse analysis from a linguistic perspective. Alvi and Baseer (2011) extracted insights from Obama’s keynote address at the 2004 DNC by counting occurrences of pronouns in 1st person singular, 2nd person, and 1st person plural. The use of pronouns indicates how the speaker wanted to be viewed by the audience. For example, the speaker used the first person singular, “I”, to emphasize personal contribution while the first-person plural, “we”, to share responsibility, build rapport, and give a sense of inclusiveness. Selby (2013) analyzed Huckabee’s use of rhetoric of proportion to minimize the significance of Romney’s religious affiliation, Mormon faith, at the 2012 RNC. Rhetoric of proportion is a strategy that Huckabee used to build identity among the disparate persons or interests by opposing a common enemy, Barack Obama. Vianica and Tanto (2021) described how lexicalization and repetition was applied to Biden’s acceptance speech at the 2020 DNC. Lexicalization is about word choice to express ideology (Vianica Tanto, 2021). Biden used this strategy to demonstrate positive self-presentation and negative other-presentation, Donald Trump. Repetition, in the form of anaphora, diacope, and antithesis, was employed in Biden’s speech to emphasize a particular point or to make it memorable for the audience. By doing so, Biden gave himself a stronger positive representation.

The dataset size for these convention-speech studies is smaller than 23, much smaller than the study we undertake. Several studies analyzed an individual speaker’s speech in a specific year. One study used a dataset from 2008 conventions (Deason Gonzales, 2012). Two studies analyzed a dataset from selected speakers over time (Benoit, Blaney Pier, 2000; Vigil, 2014).

NLP, a scientific field of computer science, is a form of analysis of human languages. More specifically, it uses “computational techniques to learn, understand, and produce human language content” (Hirschberg Manning, 2015, p. 261). The convenience of accessing large quantities of textual speech data in a digital form has made applying NLP to political speeches achievable. This trend is reflected in two recent studies that used comprehensive congressional speeches across a span of over a century, with data analysis conducted in a way to detect a chronological shift of political opinions and attitudes (Ethan, Tucker, Capps, Shamir, 2020; Card, Chang, Becker, Mendelsohn, Boigt, Boustan, Abramitzky, Jurafsky, 2022).

In terms of methods major NLP research methods have been applied to relevant studies: text preprocessing, text representation, text classification, topic modeling, and sentiment analysis. In text preprocessing, tokenization, lemmatization, and part-of-speech were adopted in a study to pre-process annotated support and attack merged under the related label (Menini, Cabrio, Tonelli Villata, 2018). In text representation Guerini, Strapparava, and Stock (2008) used tf-idf to evaluate the weight of a word to a document in a corpus. Kassarnig (2016) obtained the probabilities for each 5-gram to start a speech when generating a speech. For text classification, in a linear regression model built by Anttila, Dozat, Galbraith, and Shapiro (2018), the predictors bigram informativity, mechanical stress, part of speech, and word position were found independently significant in predicting perceived stress. Interestingly, the effect from bigram informativity to perceived stress was entirely driven by verbs and function words. Card, et al. (2022) applied logistic regression to predict tone and relevance to immigration from the most frequently mentioned nationalities. Menini et al. (2018) proposed a relation classification system to predict support and attack relations between arguments. For topic modeling, Latent Dirichlet Allocation (LDA) was used for natural language generation (Kassarnig, 2016). The topic model was built by setting 53 underlying topics from 53 different debates. As a result, LDA generated a mixture of good and bad examples. For sentiment analysis, Ethan et al. (2020) tracked sentiment changes from the two parties expressed in congressional speeches. A sentiment value 0-4 was assigned from very negative to very positive. Both very positive and very negative sentiments were captured to be more common in recent years. For more method information, please see Table 1.

Table 1

*List of References for Methods on NLP in Political Speeches*

|  |  |  |  |
| --- | --- | --- | --- |
| Authors | Approach | Description | Dataset |
| Anttila et al. (2018) | Part-of-speechLinear regression  | Capture sentential prominence in the inaugural addresses of six U.S. presidents.  | The first inaugural addresses of six presidents: Carter, Reagan, Bush Sr., Clinton, and Obama in script, audio, and video |
| Card et al. (2022)  | TokenizationBag-of-wordsManual annotationsClassificationsValidity checksLogistic regressionContextual embedding | Investigate evolution of attitudes towards immigration over three phases: early (1880-1934), transitional (1935-1972), modern (1973-2020) | 140 years of US congressional and presidential speeches about immigration between 1880-2020 |
| Ficcadenti, Cerqueti, and Ausloos (2019) | Zipf-Mandelbrot Law | Explore the implicit structure of the discourse data through a rank-size analysis on a corpus  | 951 Presidential speeches by the 45 US Presidents |
| Ethan et al. (2020) | Coleman-Liau readability index StemmingSentiment analysisWord diversityWord homogeneityTotal number of wordsTopic words | Present speech differences in terms of readability, word diversity, homogeneity, length, and sentiment over years and between the two parties.  | 138 years’ USA Congressional speeches made between 1873 to 2010 |
| Guerini et al. (2008) | LemmatizationPOS analysisNamed entity recognitionSentence splittingSentiWordNetTF-IDF | Apply persuasive opinion mining to detectpolitical speech changes; Present a corpus-based approach for persuasive expression mining that relies on NLP.  | CORPS that contains political speeches tagged with audience reactions, approximately 900 speeches |
| Kassarnig (2016) | N-gramsJusteson Katz POS tag filterRecurrent neural networksLDA | Automatically generate political speeches for a desired political party | 3857 speech segments from 53 US Congressional floor debates from the year 2005 |
| Menini et al. (2018)  | Manual annotationsBag-of-wordsTokenizationLemmatizationPart-of -speechWord2Vec | Apply argumentation mining techniques for relation prediction to study political speeches in monological form | Speeches and official declarations including 881 documents issued by Nixon and Kennedy during 1960 Presidential campaign  |
| Savoy (2010)  | Lexical richnessChi-squareLog-likelihood | Determine if a given word can be used to describe a subset;Measure the association between a word and a set of documents;Measure the association between a given word and a corpus  | 95 speeches by McCain and 150 speeches by Obama from 2007-2008 |

The research gap that this study will fill will be three-fold: first, we will use a comprehensive DNC and RNC speeches dataset with about 140 hours’ worth of videos from 2008-2020. The original data source is from C-SPAN.org. Transcript accuracy was manually ensured. Second, we will identify patterns of morphosyntactic process, word categories such as singular/plural, and verb tenses such as the past, present, and future. We will compare the differences between the two parties and interpret the results. To the best of our knowledge, no similar research has been done in existing literature. Third, we will analyze the data from the speakers by their political affiliation. All speakers from the same party convention will contribute to one speech style. An analysis of a convention speech style represented by this number of speakers, 399 speakers from the Democratic and 934 speakers from the Republican Party, has not been conducted before.

# Methods

## The datasets

The main dataset of this project is a comprehensive collection of DNC and RNC speeches from 2004-2020. Because the convention is held every four years, we have five years worth of dataset: 2004, 2008, 2012, 2016, and 2020. Our data come from rev.com and C-SPAN.org. The textual data were entered to a SQLite database with 3470 rows and 9 columns including year, party, day, speaker, speaker count, time, text, text length, and the source of text, as shown in an example in Figure 1.

Figure 1.An Overview of the Convention Speech Dataset

A typical convention goes on for four nights. All conventions were held in person except the DNC in 2020, which was held virtually because of the pandemic. The RNC in 2020 took place in person, but with a reduced number of delegates. Descriptive statistics from the dataset shown in Table 2 show how the number of speakers and speeches vary over the years. The Democratic Party overall has more speakers, speeches, total tokens, and unique tokens than the Republican Party. However, the Republican Party has longer average token length and more lexical diversity.

Table 2

*An Overview of Descriptive Statistics in Convention Speeches*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Party | Distinct Speakers | Distinct Speeches | Tokens | Uniquetokens | Average token length | Lexical diversity |
| 2020 | Democratic | 334 | 1258 | 42047 | 9416 | 6.02 | 0.22 |
| Republican | 133 | 732 | 47228 | 10683 | 6.28 | 0.23 |
| 2016 | Democratic | 200 | 221 | 61064 | 11758 | 6.15 | 0.19 |
| Republican | 68 | 76 | 33613 | 8560 | 6.2 | 0.25 |
| 2012 | Democratic | 110 | 115 | 44434 | 9071 | 6.2 | 0.2 |
| Republican | 82 | 91 | 34822 | 8501 | 6.2 | 0.24 |
| 2008 | Democratic | 110 | 113 | 43720 | 8868 | 6.16 | 0.2 |
| Republican | 55 | 61 | 26019 | 7056 | 6.22 | 0.27 |
| 2004 | Democratic | 180 | 184 | 63505 | 11206 | 6.2 | 0.18 |
| Republican | 61 | 68 | 23508 | 6842 | 6.28 | 0.29 |

An extended dataset we use for permutation testing in this study is 1,038 presidential speeches from 1789 to 2021, from George Washington to Joe Biden. These speeches were delivered by 45 U.S. Presidents, among whom 445 speeches were from 19 Republican Presidents and 513 speeches from 16 Democratic Presidents. The other ten non-Republican or non-Democratic Presidents from 1789 to 1853 are: No party (George Washington), Federalist (John Adams), Democratic-Republican (Thomas Jefferson, James Madison, James Monroe, John Quincy Adams), and Whig (William Henry Harrison, John Tyler, Zachary Taylor, Millard Fillmore). The data source is https://millercenter.org/the-presidency/presidential-speeches. Table 3 gives us an overview of this dataset. As shown in this table, Presidents Johnson, Reagan, and Obama gave the greatest number of speeches. Presidents in early years such as Presidents Garfield, Harrison, Taylor, and Adams have the most lexical diversity.

Table 3

*An Overview of Descriptive Statistics in Presidential Speeches*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| President | Party | Speeches | Total tokens | Unique tokens | Avg tokenlength | Lexical diversity |
| Lyndon B. Johnson | Democratic | 71 | 122709 | 17258 | 6.59 | 0.14 |
| Ronald Reagan | Republican | 60 | 113101 | 19722 | 6.51 | 0.17 |
| Barack Obama | Democratic | 50 | 109360 | 16719 | 6.33 | 0.15 |
| Franklin D. Roosevelt | Democratic | 49 | 65908 | 13357 | 6.61 | 0.2 |
| John F. Kennedy | Democratic | 45 | 81724 | 15455 | 6.67 | 0.19 |
| Donald Trump | Republican | 43 | 116175 | 18563 | 6.19 | 0.16 |
| George W. Bush | Republican | 40 | 60143 | 11897 | 6.56 | 0.2 |
| Bill Clinton | Democratic | 39 | 79556 | 13457 | 6.3 | 0.17 |
| Woodrow Wilson | Democratic | 33 | 40540 | 9539 | 6.85 | 0.24 |
| Ulysses S. Grant | Republican | 32 | 51138 | 11228 | 7.28 | 0.22 |
| Andrew Johnson | Democratic | 31 | 48035 | 10251 | 7.23 | 0.21 |
| Herbert Hoover | Republican | 30 | 45461 | 9605 | 7.23 | 0.21 |
| Grover Cleveland | Democratic | 30 | 75775 | 14718 | 7.38 | 0.19 |
| Andrew Jackson | Democratic | 26 | 73703 | 12094 | 7.3 | 0.16 |
| James K. Polk | Democratic | 25 | 50140 | 9101 | 7.19 | 0.18 |
| Thomas Jefferson | Democratic-Republican | 24 | 19119 | 6273 | 7.09 | 0.33 |
| Richard M. Nixon | Republican | 23 | 32940 | 7991 | 6.57 | 0.24 |
| Benjamin Harrison | Republican | 23 | 68665 | 10716 | 7.18 | 0.16 |
| George H. W. Bush | Republican | 23 | 52109 | 10962 | 6.36 | 0.21 |
| James Madison | Democratic-Republican | 22 | 17141 | 5497 | 7.42 | 0.32 |
| Jimmy Carter | Democratic | 22 | 52122 | 10541 | 6.6 | 0.2 |
| Theodore Roosevelt | Republican | 22 | 97884 | 17258 | 7.04 | 0.18 |
| George Washington  | No party | 21 | 15166 | 5788 | 7.38 | 0.38 |
| Joe Biden | Democratic | 19 | 42025 | 10445 | 6.35 | 0.25 |
| Harry S. Truman  | Democratic | 19 | 17882 | 5787 | 6.52 | 0.32 |
| John Tyler | Whig | 18 | 33409 | 8097 | 7.22 | 0.24 |
| Warren G. Harding | Republican | 18 | 14684 | 5366 | 7.31 | 0.37 |
| Rutherford B. Hayes  | Republican | 16 | 33337 | 7791 | 7.38 | 0.23 |
| Abraham Lincoln | Republican | 15 | 50655 | 10393 | 6.89 | 0.21 |
| Franklin Pierce  | Democratic | 15 | 30958 | 8083 | 7.48 | 0.26 |
| James Buchanan  | Democratic | 14 | 39441 | 8956 | 7.26 | 0.23 |
| Gerald Ford  | Republican | 14 | 21414 | 6731 | 6.87 | 0.31 |
| William McKinley | Republican | 14 | 47209 | 11873 | 7.37 | 0.25 |
| William Taft | Republican | 12 | 60238 | 11890 | 7.38 | 0.2 |
| Calvin Coolidge | Republican | 12 | 36780 | 9747 | 7.36 | 0.27 |
| Chester A. Arthur | Republican | 11 | 25.62 | 7735 | 7.47 | 0.31 |
| James Monroe | Democratic-Republican | 10 | 23210 | 5834 | 7.27 | 0.25 |
| Martin Van Buren | Democratic | 10 | 30922 | 7623 | 7.4 | 0.25 |
| John Adams | Federalist | 9 | 7020 | 3201 | 7.42 | 0.46 |
| John Quincy Adams | Democratic-Republican | 9 | 17302 | 5690 | 7.48 | 0.33 |
| Millard Fillmore | Whig | 7 | 18925 | 6253 | 7.33 | 0.33 |
| Dwight D. Eisenhower | Republican | 6 | 9345 | 4241 | 6.83 | 0.45 |
| Zachary Taylor  | Whig | 4 | 5412 | 2510 | 7.43 | 0.46 |
| James A. Garfield | Republican | 1 | 1433 | 994 | 7.25 | 0.69 |
| William Harrison  | Whig | 1 | 3873 | 2051 | 7.36 | 0.53 |

## The research approaches

### Topic modeling

The first research approach applied to this study is topic modeling. Topic modeling is a machine learning and NLP technique for determining the topics present in a text. It is a probabilistic model used to understand a topic as a theme or underlying meaning cluster with related ideas represented in text. We use Python and its packages spaCy, Gensim, and scikit-learn’s machine learning methods in our text analysis. The package spaCy “describes itself as Industrial Strength Natural Language Processing” (Srinivasa-Desikan, 2018, p.35). spaCy was used to preprocess data. Gensim, on the other hand, is used to vectorize text and perform transformations. To perform the topic modeling we use a method called Latent Dirichlet allocation (LDA), an unsupervised machine learning clustering technique. Unsupervised approaches are useful when we do not have labeled corpora ready for classification, as is the case with this text data. LDA has two advantages. First, LDA brings structure to otherwise unstructured text data (Kumar, 2018). Second, LDA models allow topic fussiness because topics are not required to be distinct (Bengfort, Bilbro, and Ojeda, 2018). In our study, we apply LDA to each year of each party’s convention speeches. Therefore, we have ten independent groups.

We use topic coherence to evaluate topic modeling. Because topic modeling does not guarantee interpretable results, topic coherence is useful to distinguish between good and bad topics (Kumar, 2018). A coherence score measures how interpretable the topics are to humans and how similar the highest probability words in a topic are to each other. It is one of the main techniques used to estimate the number of topics, but it works best when interacting with other variables. In our study we use two variables, average topic overlap and topic coherence score, to determine the optimal number of topics in each group. Ideally, we expect our topics to be low in average topic overlap and high in topic coherence.

The procedures of topic modeling are:

* Data preprocessing
	+ Cleaning
	+ Tokenization
	+ Lemmatization
* Train LDA model
* Get descriptive statistics
* Compute probabilities for each category
* Identify the optimal number of topics in each group
* Determine the topics
* Visualize the topics

### Permutation test

The linguistic theoretical framework for our permutation testing originates from three dimensions of deixis: time (*t*), space (*s*), and modality (*m*), as shown in Figure 2 (Chilton, 2004). We borrow the chart to demonstrate the connections between *t* and *s*. Tense and pronouns perform deictic functions (Chilton, 2004). At the deictic center is the Self, I or we, and the present, now. The two ends of t represent the past and the future. *t* and *s* interact in a way that time works through space, “relative distance to or from self, events, which carry a time of happening as part of their conceptualization” (Chilton, 2004, p. 59).

Figure 2. Dimensions of Deixis (Chilton, 2004, p.58)

The second research approach permutation test is a non-parametric test procedure we use to test the null hypothesis that the two parties have the same distribution in terms of tense and first-person pronoun uses. Under the null hypothesis we assume permutations from each party are equally likely. For a two-sided test, we define the alternative hypothesis that the two parties are different. Therefore, our null and alternative hypotheses are:

The null hypotheses

* + Party is independent of intensity of first person pronoun usage.
	+ Party is independent of intensity of future tense verb usage.

The alternative hypotheses

* + Republicans are more likely to use the first person singular than Democrats.
	+ Democrats are more likely to use the future tense than Republicans.

The procedures of permutation test are:

* Create a dataset with two new tables for Convention and Presidential speeches. For each new table, we first extract five columns from the existing datasets: year, party, speaker, text, file, then add six new columns: the number of verbs in past and future tenses, the number of verbs, the number of first person singular pronouns (“I”), the number of first person plural pronouns (“we”), and the number of sentences.
* Calculate the ratio of verbs in past/future tenses to the number of verbs and the ratio of “I” and “we” to the number of sentences.
* Create permuted data sets (scramble party column), calculate stat, store.
* Compare actual to permuted values to assess statistical significance under the null hypothesis.

# Results and discussions

## Topic modeling

Topic modeling generates topics for each year of each party. As mentioned earlier, we use the interactions between two variables, average topic overlap and topic coherence, to determine the optimal number of topics. As a result, we generate ten charts (one for each party each year) to visualize the interactive process. Figure 3 is an example of the visual from 2020 RNC speeches. The ideal number of topics (five) is selected because of its combination of low average topic overlap and high topic coherence. A general trend of average topic overlap is that the more topics we have, the more unlikely the topics overlap. However, we can’t have an unlimited number of topics, so we consider the coherence score. There is always the highest coherence score that happens with relatively low average topic overlap, which is the point selected.

For more coherence and topic overlap visuals, please see Appendix B.

Figure 3. Average Topic Overlap, Topic Coherence, and Ideal Number of Topics from 2020 RNC

Table 4 shows specific topics discovered for each party in each year. The smallest number of topics was at the DNC in 2020 and the largest number was at the DNC in 2016. Each topic was supported by a list of top probability words, in the order from high to low probability. We are not able to give a single topic word to represent every party in every year, because in some cases the semantic similarity is not close enough and no coherent topic emerges. In this case, we use N/A to represent that party and year. To access all probability words of each topic, please see Appendices C and D for the Republican and Democratic Parties, respectively.

As we look closer to the results, we identified similar and different topics each year. In 2004, both parties shared a common topic related to national security. The likely reason is the September 11 Attacks that happened in 2001. The Republican Party used top probability words such as “war”, “freedom”, “terrorist”, “terrorism”, “attack”, “weapon”, and “enemy”, whereas the Democratic Party was more likely to use the words “safe”, “security”, “military”, “weapon”, “protect”, “terrorist”, “terror”, and “courage”. . The words from the Republican Party take a more proactive stance since words like “war”, “enemy”, and “attack” have a more confrontational meaning, while the Democratic Party tend to be more defensive by using words such as “security”, “protect”, and “courage”. In 2008, there seemed to be a shift to the economy as a focus. Both parties talked about “oil”, “economy”, and “business”. It seemed that the Democratic Party was more interested in green energy. The Republican Party brought about the idea of challenges and changes. In 2012, it was interesting to see that although the Democratic Party presented very clear and distinct topics, it was hard to identify quality topics for the Republican Party. One possible explanation is that 2012 is the year Obama sought his second term; he had a good chance of being reelected. 2016 the name “Trump” showed up as a topic in the Democratic Party, for obvious reasons. This same year, the Republican Party presented very clear topics. One unique topic was border associated with words like “terrorist”, “terrorism”, “fail”, and “immigration”, consistent with the Republican Party’s conservative political stance towards immigration. In 2020, the Republican Party had a topic “Trump” associated with words like “America”, “nation”, “job”, “nation”, and “country”. Another new topic for Republicans that year was “drug”, associated with “addiction”, “impact”, “struggle”, and “crisis”. The Democratic Party, on the other hand, had a topics associated with words like “job”, “worker”, “family”, “healthcare”, “Trump”, “climate”, “world”, “democracy”, and “economy”, reflecting the Party’s efforts to encourage people to vote.

Table 4

*Topics from 2004 to 2020*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Party | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 |
| 2020 | Democratic | N/A | Family | Vote | President |  |  |  |  |
| Republican | Child | Trump | N/A | Nation | Drug |  |  |  |
| 2016 | Democratic | N/A | Election | Trump | N/A | Nation | Party | N/A | Military |
| Republican | Party | Business | Challenge | Bill | Freedom | Border | Father |  |
| 2012 | Democratic | Business | Veteran | Energy | Industry | Dream | Medicare | Leadership |  |
| Republican | School | N/A | Woman | N/A | N/A |  |  |  |
| 2008 | Democratic | Energy | Woman | War | Economy | Education |  |  |  |
| Republican | Change | Business | Man | Service | Support |  |  |  |
| 2004 | Democratic | Healthcare | Party | Dream | Security | School |  |  |  |
| Republican | State | Child | N/A | Tax | War | Worker |  |  |

Over time the two parties have both some topics that have shifted and some that are persistent. The Republicans shifted their topics from war to business and school, then to nationalism related to the border, immigration, and job security. The Democrats shifted from national security, to the economy and energy, then to elections and voting. The Republican Party persistently has had religious topics such as father, son, economic topics such as business and taxes, and national interest topics related to freedom and nation. The Democratic Party has consistently had the future-oriented topic, “dream”, an energy topic such as “green energy”, and a healthcare topic such as “Medicare”.

Below is a snapshot from interactive topic model visualizations. On the left, the number of bubbles represents the number of topics. The size of each bubble represents the percentage of tokens. The more the bubble is, the more weight it carries in each group. On the right is the list of top-30 most relevant terms for the selected topic as indicated by Red.

Figure 4. A Snapshot from Interactive Democratic 2004 Topic Model Visualization with *pyLDAvis*

## Permutation test

Our permutation test started with the building of two tables , one from convention speeches and one from the presidential speeches. To create the two tables, we first extracted five columns:  year, party, speaker, text, and file, then added six newly created columns: the number of verbs in past and future tenses, the number of verbs, the number of first person singular pronouns (“I”), the number of first person plural pronouns (“we”), and the number of sentences. Below is a sample of our new table for presidential speeches. Another table for convention speeches has the same structure.

Figure 5.An Overview of the Presidential Speech Table****

Table 5 shows the descriptive statistics of the six new columns. Overall the Presidential speeches are longer d therefore end up having higher medians for all six variables than the convention speeches. In convention speeches, the minimum verb occurrence 0 is likely from very short texts. It is currently not clear why presidential speeches include a speech with only two sentences.

Table 5

*Descriptive Statistics of Newly Created Columns by Count*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Speech  | Statistics | Future tense | Past tense | Verb | Plural | Singular | Sentence |
| ConventionSpeeches | Min. | 0 | 0 | 0 | 0 | 0 | 1 |
| 1st Qu. | 0 | 0 | 2 | 0 | 0 | 1 |
| Median | 0 | 1 | 10.5 | 0 | 1 | 5 |
| Mean | 0.34 | 4.82 | 27.43 | 1.65 | 3.17 | 17.17 |
| 3rd Qu. | 0 | 5 | 27 | 1 | 3 | 15.75 |
| Max | 29 | 294 | 829 | 134 | 148 | 398 |
| SD | 1.27 | 11.62 | 58.75 | 6.87 | 7.74 | 33.7 |
| PresidentialSpeeches | Min. | 0 | 0 | 10 | 0 | 0 | 2 |
| 1st Qu. | 2 | 9 | 129 | 5 | 7 | 41 |
| Median | 5 | 25 | 299 | 21 | 19 | 113.5 |
| Mean | 6.87 | 42.05 | 442 | 40.83 | 36.51 | 167.7 |
| 3rd Qu. | 10 | 53 | 579 | 58 | 43.75 | 221 |
| Max | 67 | 636 | 4137 | 711 | 640 | 2119 |
| SD | 7.44 | 56.29 | 464.6 | 58.43 | 56.1 | 191.23 |

From the permutation tests, we find that p-value is smaller than 0.05 for past tense and singular pronoun usage in convention speeches and past tense in Presidential speeches. That is to say, the Republican Party is more likely to use past tense and singular first person pronouns than the Democratic Party in their convention speeches, and to use past tense in their Presidential speeches. No statistically significant differences are found in terms of future tense and plural pronoun usage between the two parties in either corpora.

In political speeches the choice of pronouns is related to the speakers’ attitude, social status, gender, and motivation (Alavidze, 2017). Personal pronouns are powerful strategies that politicians use to achieve their goals. By using deictic words of their preference, politicians deliver their intention to people. The relative usage of singular first person pronouns versus plural pronouns reflects levels of individualism. We interpret the results to indicate that the Republican party values the “independent self” over the “interdependent self” to a larger degree than the Democratic Party. The Republican Party tends to speak from their individual perspective, showg authority, personal responsibility, commitment, and involvement. The Democratic Party, on the other hand, prefers to create involvement from the audience, share the responsibility, and give a sense of collectivity.

Our permutation test results for past tense usage in both corpora are consistent with the existing literature regarding the two parties’ perceptions on the future and the past. Robinson, Cassidy, Boyd, and Fetterman (2015) stated that the Republican Party favors certainty and values tradition whereas the Democratic party favors change. Certainty and tradition suggest a preference to the past while change suggests a preference to the future. Conservatives, to a great extent, endorse the values of conservation to maintain and preserve traditions. The usage of past tense suggests a pattern of thinking that the conservatives value security and certainty more than openness to change.

# Conclusion

In this study, we analyze five years of convention speeches fby the Democratic and Republican Parties from 2004 to 2020 and 1,038 presidential speeches from 45 U.S. Presidents from 1789 ( George Washington) to 2021 (Joe Biden). We use two research approaches: topic modeling and permutation test, to analyze the textual data. Our topic modeling identifies topics that gain or lose favor over time and topics that consistently reflect core values of the two parties. Our permutation test analysis shows statistically significant differences in past tense usage between the two parties in two corpora and in singular pronoun usage in convention speeches.

In the present work we applied topic modeling to the convention speeches because the Presidential speeches have been going on for over two hundred years. It is ideal to analyze this corpus from phases of early, transitional, and modern in the future. Moreover, we have limited ourselves to computational morphosyntactic analysis of English political speeches. In the long run, we will conduct a study to compare political speeches between languages using NLP techniques.

**Acknowledgements.** My sincere thanks go to Dr. John Chandler for letting me use his datasets for my capstone project and for his expertise supervising this project.

# References

Alvi, S. D., Baseer, A. (2011). An analysis of Barack Obama’s speech: keynote address at the 2004 Democratic National Convention. *Language in India*, *11*(10), 310–335.

Alavidze, M. (2017). The use of pronouns in political speeches. *International Journal of Arts Sciences,*9(4), 349–356.

Anttila, A., Dozat, T., Galbraith, D., Shapiro, N. (2018). Sentence stress in presidential speeches. *LingBuzz*.

Benoit, W. L., Blaney, J. R., Pier, P. M. (2000). Acclaiming, attacking, and defending: A functional analysis of U.S. nominating convention keynote speeches. *Political Communication*, *17*(1), 61–84. <https://doi.org/10.1080/105846000198512>.

Bengfort, B., Bilbro, R., Ojeda, T. (2018). *Applied text analysis with Python enabling language-aware data products with machine learning* (1st ed.). O’Reilly.

Card, D., Chang, S., Becker, C., Mendelsohn, J., Voigt, R., Boustan, L., Abramitzky, R., Jurafsky, D. (2022). Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences - PNAS*, *119*(31), 1–e2120510119. <https://doi.org/10.1073/pnas.2120510119>.

Charteries-Black, J. (2018). *Analysing political speeches: rhetoric, discourse and metaphor* (2nd ed.). London: Palgrave.

Chilton, P. (2004). *Analysing political discourse: theory and practice*. London: Routledge. https://doi.org/10.4324/9780203561218.

Deason, G., Gonzales, M. H. (2012). Moral politics in the 2008 presidential convention acceptance speeches. *Basic and Applied Social Psychology*, *34*(3), 254–268. <https://doi.org/10.1080/01973533.2012.674450>.

Ficcadenti, V., Cerqueti, R., Ausloos, M. (2019). A joint text mining-rank size investigation of the rhetoric structures of the US Presidents’ speeches. *Expert Systems with Applications*, *123*, 127–142. <https://doi.org/10.1016/j.eswa.2018.12.049>.

Frank, D. A., McPhail, M. L. (2005). Barack Obama’s address to the 2004 Democratic National Convention: trauma, compromise, consilience, and the (im)possibility of racial reconciliation. *Rhetoric Public Affairs*, *8*(4), 571–593. <https://doi.org/10.1353/rap.2006.0006>.

Guerini, M., Strapparava, C., Stock, O. (2008). CORPS: A Corpus of tagged political speeches for persuasive communication processing. *Journal of Information Technology Politics*, *5*(1), 19–32. <https://doi.org/10.1080/19331680802149616>.

Hirschberg, J., Manning, C. D. (2015). Advances in natural language processing. *Science (American Association for the Advancement of Science)*, *349*(6245), 261–266. <https://doi.org/10.1126/science.aaa8685>.

Holbert, R. L., Hardy, B. W., LaMarre, H. L. (2017). A Normative Assessment of 2016 Political Convention Speech Exposure: Perceived Political Threats and Anticipated General Election Legitimacy. *The American Behavioral Scientist (Beverly Hills)*, *61*(4), 379–400. <https://doi.org/10.1177/0002764217693275>.

Kashima, Y., Kashima, E. S. (2003). Individualism, GNP, climate, and pronoun drop: Is individualism determined by affluence and climate, or does language use play a role? *Journal of Cross-Cultural Psychology*, 34, 125-134.

Kassarnig, V. (2016). *Political speech generation*. <https://doi.org/10.48550/arxiv.1601.03313>.

Kendall, K. E. (2017). Why did Khizr Khan’s speech at the 2016 Democratic National Convention go viral? Personifying collective values in an epideictic speech. *The American Behavioral Scientist (Beverly Hills)*, *61*(6), 611–623. <https://doi.org/10.1177/0002764217723044>.

Kumar, K. (2018). *Evaluation of topic modeling: topic coherence*. Retrieved from <https://datascienceplus.com/evaluation-of-topic-modeling-topic-coherence/>.

Menini, S., Cabrio, E., Tonelli, S., Villata, S. (2018). Never retreat, never retract: argumentation analysis for political speeches. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). <https://doi.org/10.1609/aaai.v32i1.11920>.

Robinson, M.D., Cassidy, D. M., Boyd, R. L., Fetterman, A. K. (2015). The politics of time: Conservatives differentially reference the past and liberals differentially reference the future: Conservatives differentially reference the past. *Journal of Applied Social Psychology*, 45(7), 391–399. <https://doi.org/10.1111/jasp.12306>.

Savoy, J. (2010). Lexical analysis of US political speeches. *Journal of Quantitative Linguistics*, *17*(2), 123–141. <https://doi.org/10.1080/09296171003643205>.

Schaffner, C. (1996). Editorial: ‘Political speeches and discourse analysis’, *Current issues in Language Society*, 3, (3).

Selby, G. S. (2013). “Where Mitt Romney Takes His Family to Church”: Mike Huckabee’s GOP Convention speech, the “Mormon Hurdle,” and the rhetoric of proportion. *Rhetoric Public Affairs*, *16*(2), 385–400. <https://doi.org/10.14321/rhetpublaffa.16.2.0385>.

Sheckels, T. F. (2017). The Keynotes of the 2016 political conventions: the death of a genre? *The American Behavioral Scientist (Beverly Hills)*, *61*(4), 401–413. <https://doi.org/10.1177/0002764217693276>.

Srinivasa-Desikan, B. (2018). *Natural Language Processing and Computational Linguistics: a practical guide to text analysis with Python, Gensim, spaCy, and Keras* (1st edition). Birmingham-Mumbai: Packt.

Tucker, C. E., Capps, C. J., Shamir, L. (2020). A data science approach to 138 years of congressional speeches. *Heliyon*, *6*(8), e04417–e04417. <https://doi.org/10.1016/j.heliyon.2020.e04417>.

Uz, I.  (2014). Individualism and first person pronoun use in written texts across languages. *Journal of Cross-Cultural Psychology*, *45*(10), 1671–1678. <https://doi.org/10.1177/0022022114550481>.

Vianica, I. Tanto, T. (2021). Representation of the self and other in Joe Biden’s Democratic National Convention speech. *Insaniyat (Online)*, *6*(1), 57–69. <https://doi.org/10.15408/insaniyat.v6i1.20575>.

Vigil, T. R. (2014). Feminine views in the feminine style: convention speeches by presidential nominees’ spouses. *The Southern Communication Journal*, *79*(4), 327–346. <https://doi.org/10.1080/1041794X.2014.916339>.

# Appendix A

Topic modeling: interactive topic modeling visuals

 Please see attached html files.

# Appendix B

Topic modeling: optimal number of topics visuals















# Appendix C

Topic modeling: top probability words for the Republican Party

|  |  |  |
| --- | --- | --- |
| Year | Topic | Top probability words |
| 2020 | Child | 0.023\*have + 0.019\*do + 0.018\*people + 0.017\*make + 0.017\*child + 0.016\*country + 0.015\*know + 0.014\*want + 0.014\*so + 0.014\*year + 0.014\*family + 0.013\*work + 0.013\*president + 0.013\*more + 0.011\*many + 0.010\*just + 0.010\*time + 0.010\*day + 0.010\*’ + 0.009\*say |
| Trump | 0.015\*year + 0.015\*american + 0.015\*trump + 0.014\*job + 0.013\*take + 0.012\*economy + 0.011\*law + 0.010\*back + 0.009\*woman + 0.009\*president + 0.008\*first + 0.008\*again + 0.008\*more + 0.008\*cut + 0.008\*radical + 0.008\*deal + 0.007\*vote + 0.007\*world+ 0.007\*support + 0.007\*city |
| N/A | 0.049\*news + 0.035\*week + 0.029\*get + 0.027\*stay + 0.026\*read + 0.025\*kind + 0.024\*night + 0.023\*most + 0.023\*’ + 0.023\*form + 0.023\*important + 0.023\*complete + 0.019\*company + 0.015\*want + 0.015\*ago + 0.014\*business + 0.014\*national + 0.012\*work + 0.011\*bless + 0.011\*press |
| Nation | 0.025\*country + 0.024\*american + 0.019\*nation + 0.017\*freedom + 0.015\*fight + 0.015\*continue + 0.014\*today + 0.014\*dream + 0.012\*people + 0.012\*first + 0.012\*history + 0.011\*great + 0.010\*woman + 0.010\*believe + 0.010\*life + 0.010\*opportunity + 0.009\*lady + 0.009\*citizen + 0.009\*father + 0.008\*right |
| Drug | 0.034\*addiction + 0.029\*help + 0.024\*drug + 0.014\*impact + 0.011\*focus + 0.010\*resource + 0.010\*save + 0.010\*even + 0.009\*need + 0.009\*administration + 0.009\*allow + 0.008\*life + 0.008\*industry + 0.008\*ask + 0.008\*amazing + 0.008\*provide + 0.008\*crisis + 0.008\*struggle + 0.007\*organization + 0.007\*nation |
| 2016 | Party | 0.024\*party + 0.023\*’ + 0.013\*there + 0.012\*guy + 0.012\*start + 0.011\*’s + 0.010\*still + 0.010\*idea + 0.010\*’m + 0.009\*look + 0.009\*win + 0.008\*break + 0.008\*much + 0.008\*move + 0.008\*team + 0.008\*turn + 0.007\*hear + 0.007\*grow + 0.007\*free + 0.007\*ready |
| Business | 0.047\*business + 0.014\*small + 0.012\*employee + 0.012\*future + 0.011\*understand + 0.009\*trade + 0.009\*hear + 0.009\*republican + 0.009\*opportunity + 0.009\*start + 0.008\*taxis + 0.008\*high + 0.008\*pay + 0.008\*policy + 0.008\*economy + 0.008\*regulation + 0.008\*past + 0.008\*enough + 0.007\*create + 0.007\*plan |
| Challenge | 0.018\*challenge + 0.015\*trust + 0.014\*community + 0.012\*help + 0.012\*leadership + 0.011\*turn + 0.011\*understand + 0.010\*party + 0.009\*elect + 0.009\*double + 0.009\*’s + 0.009\*cost + 0.008\*faith + 0.008\*business + 0.008\*together + 0.008\*’ + 0.008\*policy + 0.008\*hope + 0.008\*home + 0.007\*promise |
| Bill | 0.021\*’s + 0.017\*bill + 0.016\*lie + 0.015\*’ + 0.012\*fact + 0.011\*pass + 0.010\*lead + 0.010\*security + 0.009\*last + 0.009\*keep + 0.008\*question + 0.008\*home + 0.008\*less + 0.008\*chief + 0.008\*deal + 0.008\*history + 0.008\*serve + 0.007\*together + 0.007\*office + 0.007\*’m |
| Freedom | 0.026\*freedom + 0.017\*leadership + 0.017\*vote + 0.013\*own + 0.012\*commander + 0.012\*enemy + 0.011\*’ + 0.010\*chief + 0.010\*fail + 0.010\*lead + 0.010\*difference + 0.009\*elect + 0.009\*last + 0.008\*war + 0.008\*trust + 0.008\*live + 0.008\*home + 0.008\*word + 0.008\*win + 0.007\*night |
| Border | 0.011\*border + 0.009\*deserve + 0.009\*terrorist + 0.008\*kill + 0.008\*ever + 0.008\*terrorism + 0.008\*immigration + 0.008\*respect + 0.008\*new + 0.008\*fail + 0.008\*radical + 0.008\*deal + 0.007\*strong + 0.007\*change + 0.007\*trade + 0.007\*citizen + 0.007\*lose + 0.007\*home + 0.007\*system + 0.006\*long |
| Father | 0.037\*father + 0.011\*run + 0.011\*long + 0.009\*together + 0.009\*hard + 0.008\*son + 0.008\*own + 0.008\*ever + 0.008\*look + 0.008\*matter + 0.008\*real + 0.007\*much + 0.007\*opportunity + 0.007\*generation +0.007\*start + 0.007\*help + 0.007\*talk + 0.007\*always + 0.007\*live + 0.007\*city |
| 2012 | School | 0.020\*school + 0.019\*student + 0.016\*leadership + 0.014\*stand + 0.013\*parent + 0.013\*promise + 0.013\*olympic + 0.012\*teacher + 0.010\*choice + 0.010\*keep + 0.010\*hand + 0.010\*reform + 0.009\*path + 0.009\*education + 0.009\*courage + 0.009\*election + 0.009\*high + 0.009\*opportunity + 0.009\*generation + 0.009\*kid |
| N/A | 0.019\*cheer + 0.018\*applause + 0.014\*new + 0.014\*right + 0.012\*think + 0.011\*debt + 0.010\*own + 0.009\*ask + 0.008\*administration + 0.008\*change + 0.008\*generation + 0.008\*barack + 0.008\*turn + 0.007\*woman + 0.007\*even + 0.007\*hope + 0.007\*money + 0.007\*there + 0.007\*still + 0.006\*obama |
| Woman | 0.013\*woman + 0.013\*convention + 0.013\*honor + 0.012\*election + 0.012\*friend + 0.011\*care + 0.010\*city + 0.010\*right + 0.010\*welcome + 0.009\*hear + 0.009\*fight + 0.009\*leadership + 0.008\*serve + 0.008\*stand + 0.008\*support + 0.008\*administration + 0.007\*tonight + 0.007\*trust + 0.007\*opportunity + 0.007\*important |
| N/A | 0.022\*never + 0.020\*cheer + 0.018\*applause + 0.014\*barack + 0.012\*taxis + 0.012\*freedom + 0.011\*tax + 0.011\*budget + 0.011\*obama + 0.010\*energy + 0.010\*middle + 0.009\*policy + 0.009\*owner + 0.009\*bad + 0.009\*spirit + 0.008\*cut + 0.008\*class + 0.008\*also + 0.008\*fail + 0.008\*too |
| N/A | 0.014\*story + 0.012\*help + 0.012\*never + 0.011\*love + 0.010\*stand + 0.010\*other + 0.010\*become + 0.009\*most + 0.009\*always + 0.008\*think + 0.008\*live + 0.008\*new + 0.008\*tonight + 0.007\*home + 0.007\*challenge + 0.007\*freedom + 0.007\*company + 0.007\*son + 0.007\*parent + 0.007\*free |
| 2008 | Change | 0.021\*change + 0.017\*let + 0.015\*fight + 0.013\*friend + 0.012\*job + 0.012\*lead + 0.012\*win + 0.010\*keep + 0.010\*governor + 0.010\*freedom + 0.009\*strong + 0.009\*lose + 0.009\*big + 0.009\*vote + 0.008\*school + 0.008\*free + 0.008\*republican + 0.008\*plan + 0.008\*experience + 0.007\*thing |
| Business | 0.025\*business + 0.022\*health + 0.020\*care + 0.019\*job + 0.018\*small + 0.016\*energy + 0.015\*prosperity + 0.014\*cost + 0.014\*create + 0.013\*taxis + 0.013\*high + 0.012\*father + 0.011\*then + 0.011\*choice + 0.010\*applause + 0.010\*tax + 0.010\*many + 0.009\*money + 0.009\*economy + 0.009\*individual |
| Man | 0.034\*man + 0.019\*fight + 0.017\*be + 0.015\*reform + 0.015\*oil + 0.015\*election + 0.015\*small + 0.014\*thing + 0.013\*special + 0.012\*fellow + 0.012\*energy + 0.012\*bring + 0.012\*taxis + 0.011\*raise + 0.011\*nominee + 0.011\*town + 0.011\*interest + 0.011\*office + 0.010\*tax + 0.010\*too |
| Service | 0.017\*tonight + 0.017\*service + 0.016\*man + 0.015\*together + 0.012\*ask + 0.012\*love + 0.012\*look + 0.011\*woman + 0.011\*character + 0.011\*history + 0.011\*live + 0.010\*spirit + 0.010\*bring + 0.009\*mother + 0.009\*fellow + 0.008\*hand + 0.008\*heart + 0.008\*challenge + 0.008\*hope + 0.008\*thing |
| Support | 0.012\*support + 0.012\*office + 0.012\*vice + 0.010\*question + 0.009\*lady + 0.009\*candidate + 0.009\*be + 0.009\*order + 0.008\*move + 0.006\*provide + 0.006\*part + 0.006\*nominee + 0.006\*like + 0.006\*convention + 0.006\*send + 0.004\*offer + 0.004\*mean + 0.004\*food + 0.004\*forward + 0.004\*protect |
| 2004 | State | 0.024\*state + 0.024\*governor + 0.021\*lady + 0.020\*young + 0.019\*give + 0.019\*opportunity + 0.018\*welcome + 0.018\*stand + 0.016\*convention + 0.016\*gentleman + 0.015\*serve + 0.015\*speak + 0.013\*proud + 0.013\*strong + 0.012\*party + 0.012\*support + 0.012\*heart + 0.012\*courage + 0.012\*republican + 0.011\*free |
| Child | 0.027\*child + 0.024\*life + 0.018\*school + 0.017\*first + 0.016\*just + 0.014\*education + 0.014\*believe + 0.014\*also + 0.014\*want + 0.013\*faith + 0.012\*leave + 0.012\*high + 0.011\*dream + 0.011\*hope + 0.010\*promise + 0.010\*parent + 0.010\*opportunity + 0.010\*see + 0.009\*own + 0.009\*friend |
| N/A | 0.026\*go + 0.021\*cheer + 0.020\*tell + 0.019\*then + 0.019\*applause + 0.017\*believe + 0.015\*stand + 0.015\*world + 0.014\*want + 0.014\*back + 0.012\*see + 0.012\*war + 0.011\*be + 0.011\*get + 0.010\*freedom + 0.010\*attack + 0.010\*think + 0.010\*tonight + 0.009\*soldier + 0.009\*woman |
| Tax | 0.034\*city + 0.033\*tax + 0.029\*business + 0.027\*job + 0.025\*small + 0.020\*economic + 0.020\*election + 0.018\*back + 0.017\*government + 0.016\*taxis + 0.016\*cut + 0.014\*vote + 0.014\*past + 0.013\*win + 0.013\*mean + 0.011\*high + 0.011\*republican + 0.011\*strong + 0.009\*get + 0.009\*war |
| War | 0.030\*war + 0.019\*freedom + 0.018\*world + 0.016\*terrorist + 0.015\*fight + 0.015\*see + 0.012\*stand + 0.012\*history + 0.012\*terrorism + 0.011\*attack + 0.011\*give + 0.010\*woman + 0.010\*face + 0.010\*just + 0.010\*much + 0.009\*life + 0.009\*vote + 0.009\*important + 0.009\*weapon + 0.009\*enemy |
| Worker | 0.017\*worker + 0.016\*see + 0.015\*job + 0.014\*world + 0.013\*child + 0.013\*freedom + 0.012\*many + 0.010\*home + 0.010\*terrorist + 0.010\*liberty + 0.010\*school + 0.009\*opportunity + 0.009\*tax + 0.009\*life + 0.009\*believe + 0.009\*provide + 0.009\*stand + 0.009\*woman + 0.009\*generation + 0.009\*act |

# Appendix D

Topic modeling: top probability words for the Democratic Party

|  |  |  |
| --- | --- | --- |
| Year | Topic | Top probability words |
| 2020 | N/A | 0.033\*get + 0.028\*stay + 0.028\*kind + 0.028\*’ + 0.027\*most + 0.026\*week + 0.025\*want + 0.025\*read + 0.022\*enable + 0.022\*night + 0.022\*important + 0.021\*form + 0.019\*speaker + 0.017\*ago + 0.017\*convention + 0.017\*hour + 0.016\*work + 0.014\*thank + 0.012\*national + 0.010\*democratic |
| Family | 0.042\*’ + 0.027\*go + 0.025\*know + 0.024\*have + 0.019\*get + 0.019\*do + 0.017\*family + 0.016\*see + 0.013\*make + 0.012\*say + 0.012\*just + 0.012\*life + 0.012\*so + 0.011\*big + 0.010\*people + 0.010\*well + 0.010\*tell + 0.009\*take + 0.009\*time + 0.009\*ask |
| Vote | 0.048\*vote + 0.030\*job + 0.022\*plan + 0.017\*worker + 0.014\*work + 0.013\*more + 0.013\*family + 0.013\*healthcare + 0.013\*need + 0.012\*build + 0.012\*get + 0.012\*make + 0.012\*trump + 0.012\*million + 0.012\*have + 0.011\*climate + 0.011\*pay + 0.011\*world + 0.011\*democracy + 0.011\*economy |
| President | 0.025\*president + 0.023\*woman + 0.020\*fight + 0.019\*country + 0.016\*make + 0.015\*together + 0.015\*so + 0.014\*nation + 0.012\*people + 0.011\*now + 0.011\*more + 0.011\*bring + 0.010\*love + 0.010\*work + 0.009\*good + 0.009\*american + 0.009\*black + 0.009\*let + 0.009\*come + 0.008\*right |
| 2016 | N/A | 0.001\*bridge + 0.001\*november + 0.001\*congressman + 0.001\*support + 0.001\*powerful + 0.001\*kill + 0.001\*name + 0.001\*proud + 0.001\*immigrant + 0.001\*speak + 0.001\*represent + 0.001\*road + 0.001\*act + 0.001\*service + 0.001\*movement + 0.001\*form + 0.001\*living + 0.001\*member + 0.001\*understand + 0.001\*gentleman |
| Election | 0.023\*election + 0.020\*cheer + 0.014\*much + 0.014\*understand + 0.012\*wage + 0.012\*very + 0.012\*job + 0.011\*campaign + 0.011\*health + 0.010\*justice + 0.010\*vote + 0.010\*tonight + 0.010\*platform + 0.009\*democratic + 0.009\*trump + 0.009\*bring + 0.009\*future + 0.009\*million + 0.009\*care + 0.009\*let |
| Trump | 0.011\*trump + 0.009\*job + 0.009\*tell + 0.008\*help + 0.008\*well + 0.008\*now + 0.008\*never + 0.008\*pay + 0.008\*american + 0.008\*let + 0.007\*keep + 0.007\*build + 0.007\*look + 0.007\*hard + 0.007\*care + 0.007\*kid + 0.006\*other + 0.006\*parent + 0.006\*put + 0.006\*school |
| N/A | 0.025\*throw + 0.015\*stay + 0.003\*well + 0.003\*senator + 0.002\*secretary + 0.002\*listen + 0.002\*lady + 0.002\*campaign + 0.002\*process + 0.002\*general + 0.002\*amazing + 0.002\*over + 0.002\*knock + 0.002\*area + 0.002\*citizen + 0.002\*up + 0.002\*rise + 0.002\*ground + 0.002\*begin + 0.002\*become |
| Nation | 0.032\*nation + 0.014\*gun + 0.012\*love + 0.012\*community + 0.011\*heart + 0.010\*officer + 0.009\*call + 0.009\*police + 0.009\*other + 0.008\*fear + 0.008\*democracy + 0.008\*value + 0.008\*common + 0.007\*justice + 0.007\*history + 0.007\*violence + 0.006\*join + 0.006\*stop + 0.006\*hear + 0.006\*many |
| Party | 0.021\*party + 0.018\*vote + 0.014\*win + 0.012\*state + 0.009\*proud + 0.009\*platform + 0.009\*democratic + 0.009\*voice + 0.009\*power + 0.008\*process + 0.007\*sander + 0.007\*support + 0.007\*reform + 0.007\*campaign + 0.007\*progressive + 0.006\*committee + 0.006\*organize + 0.006\*elect + 0.006\*member + 0.006\*ask |
| N/A | 0.165\*cheer + 0.017\*applause + 0.007\*hillary + 0.007\*bear + 0.006\*proud + 0.006\*continue + 0.006\*citizen + 0.006\*feel + 0.005\*candidate + 0.005\*deserve + 0.005\*part + 0.005\*never + 0.005\*show + 0.004\*introduce + 0.004\*support + 0.004\*name + 0.004\*only + 0.004\*bridge + 0.004\*movement + 0.004\*nomination |
| Military | 0.014\*world + 0.013\*veteran + 0.012\*military + 0.012\*commander + 0.012\*serve + 0.010\*isis + 0.010\*chief + 0.009\*honor + 0.007\*ready + 0.007\*trump + 0.007\*choose + 0.007\*ally + 0.007\*defeat + 0.007\*fellow + 0.007\*american + 0.006\*candidate + 0.006\*defend + 0.006\*force + 0.006\*leadership + 0.006\*man |
| 2012 | Business | 0.020\*dream + 0.019\*business + 0.015\*opportunity + 0.014\*school + 0.013\*college + 0.013\*student + 0.012\*teacher + 0.011\*small + 0.011\*start + 0.010\*love + 0.010\*nation + 0.009\*young + 0.009\*grow + 0.008\*there + 0.007\*ago + 0.007\*think + 0.007\*matter + 0.007\*invest + 0.007\*high + 0.006\*kid |
| Veteran | 0.024\*veteran + 0.017\*war + 0.015\*promise + 0.012\*military + 0.012\*serve + 0.011\*service + 0.011\*world + 0.010\*vote + 0.008\*mom + 0.008\*then + 0.008\*policy + 0.008\*leadership + 0.008\*bless + 0.007\*choice + 0.007\*ask + 0.007\*there + 0.007\*troop + 0.007\*generation + 0.007\*end + 0.007\*father |
| Energy | 0.018\*new + 0.016\*energy + 0.013\*hope + 0.011\*well + 0.010\*world + 0.010\*long + 0.009\*share + 0.009\*government + 0.009\*cut + 0.008\*choose + 0.008\*tax + 0.008\*never + 0.008\*worker + 0.007\*ask + 0.007\*still + 0.007\*choice + 0.007\*call + 0.007\*company + 0.007\*big + 0.007\*war |
| Industry | 0.016\*governor + 0.012\*industry + 0.011\*auto + 0.010\*look + 0.010\*bring + 0.010\*sector + 0.010\*private + 0.010\*call + 0.010\*create + 0.010\*worker + 0.009\*company + 0.009\*tough + 0.008\*never + 0.008\*save + 0.008\*world + 0.008\*think + 0.008\*car + 0.007\*new + 0.007\*heart + 0.007\*well |
| Dream | 0.016\*dream + 0.011\*house + 0.011\*vote + 0.009\*pass + 0.008\*act + 0.008\*democratic + 0.008\*democrat + 0.007\*cheer + 0.007\*wrong + 0.007\*republican + 0.007\*try + 0.007\*fair + 0.006\*strengthen + 0.006\*evening + 0.005\*special + 0.005\*interest + 0.005\*private + 0.005\*security + 0.005\*social + 0.005\*month |
| Medicare | 0.016\*cut + 0.014\*medicare + 0.013\*vote + 0.011\*republican + 0.010\*senior + 0.010\*cheer + 0.010\*insurance + 0.008\*tax + 0.008\*governor + 0.008\*try + 0.008\*budget + 0.008\*election + 0.008\*reform + 0.008\*law + 0.007\*put + 0.007\*too + 0.007\*promise + 0.007\*party + 0.007\*democratic + 0.006\*turn |
| Leadership | 0.023\*cheer + 0.015\*leader + 0.012\*party + 0.010\*strong + 0.010\*face + 0.010\*challenge + 0.009\*change + 0.009\*economic + 0.009\*obama + 0.009\*show + 0.009\*father + 0.008\*rule + 0.008\*inspire + 0.008\*well + 0.008\*again + 0.008\*leadership + 0.008\*lead + 0.008\*put + 0.008\*justice + 0.007\*friend |
| 2008 | Energy | 0.039\*energy + 0.022\*oil + 0.012\*create + 0.012\*tax + 0.010\*put + 0.010\*economy + 0.009\*clean + 0.009\*future + 0.008\*invest + 0.008\*generation + 0.008\*break + 0.008\*company + 0.007\*plan + 0.007\*high + 0.007\*as + 0.007\*green + 0.007\*fuel + 0.007\*own + 0.007\*renewable + 0.007\*understand |
| Woman | 0.020\*woman + 0.011\*story + 0.010\*promise + 0.009\*mother + 0.008\*young + 0.008\*college + 0.008\*mom + 0.008\*daughter + 0.008\*never + 0.008\*keep + 0.007\*fight + 0.007\*love + 0.007\*parent + 0.007\*single + 0.007\*night + 0.007\*live + 0.007\*look + 0.007\*tonight + 0.007\*applause + 0.007\*there |
| War | 0.011\*war + 0.010\*leader + 0.009\*serve + 0.009\*fight + 0.008\*elect + 0.008\*again + 0.008\*veteran + 0.008\*lead + 0.008\*honor + 0.008\*leadership + 0.008\*party + 0.008\*military + 0.008\*troop + 0.007\*keep + 0.007\*very + 0.007\*woman + 0.007\*end + 0.007\*let + 0.007\*tonight + 0.007\*never |
| Economy | 0.021\*economy + 0.019\*class + 0.016\*middle + 0.014\*lose + 0.013\*friend + 0.013\*union + 0.013\*cost + 0.011\*worker + 0.011\*tax + 0.011\*company + 0.008\*way + 0.008\*call + 0.008\*too + 0.007\*be + 0.007\*up + 0.007\*economic + 0.007\*strong + 0.007\*oil + 0.007\*mccain + 0.007\*thing |
| Education | 0.021\*education + 0.019\*school + 0.017\*business + 0.015\*future + 0.012\*economy + 0.011\*small + 0.010\*promise + 0.009\*kid + 0.009\*only + 0.008\*leave + 0.008\*look + 0.008\*think + 0.007\*high + 0.007\*student + 0.007\*town + 0.007\*race + 0.007\*plan + 0.007\*move + 0.007\*sure + 0.007\*let |
| 2004 | Healthcare | 0.037\*healthcare + 0.017\*economy + 0.017\*plan + 0.016\*city + 0.015\*cost + 0.015\*quality + 0.015\*energy + 0.014\*affordable + 0.012\*way + 0.012\*find + 0.012\*last + 0.011\*insurance + 0.011\*business + 0.010\*worker + 0.010\*friend + 0.009\*pay + 0.009\*policy + 0.009\*healthy + 0.009\*bear + 0.008\*fellow |
| Party | 0.013\*party + 0.010\*woman + 0.010\*democratic + 0.008\*vote + 0.007\*leadership + 0.007\*friend + 0.007\*convention + 0.007\*government + 0.007\*policy + 0.007\*proud + 0.007\*election + 0.007\*bring + 0.006\*generation + 0.006\*win + 0.006\*power + 0.006\*celebrate + 0.006\*protect + 0.006\*never + 0.006\*fellow + 0.005\*history |
| Dream | 0.016\*believe + 0.010\*hard + 0.009\*way + 0.009\*dream + 0.008\*live + 0.008\*young + 0.008\*never + 0.007\*father + 0.007\*tonight + 0.007\*be + 0.007\*pay + 0.006\*parent + 0.006\*still + 0.006\*back + 0.006\*even + 0.006\*woman + 0.006\*serve + 0.006\*call + 0.006\*bring + 0.005\*think |
| Security | 0.014\*safe + 0.013\*security + 0.012\*weapon + 0.011\*understand + 0.011\*military + 0.010\*city + 0.009\*police + 0.009\*service + 0.009\*keep + 0.009\*secure + 0.009\*protect + 0.008\*ask + 0.008\*terrorist + 0.008\*terror + 0.008\*courage + 0.007\*tonight + 0.007\*responsibility + 0.007\*firefighter + 0.007\*choose + 0.007\*respect |
| School | 0.048\*school + 0.032\*education + 0.025\*public + 0.019\*science + 0.017\*program + 0.016\*technology + 0.015\*support + 0.015\*teacher + 0.014\*administration + 0.014\*create + 0.013\*effort+ 0.012\*research + 0.012\*begin + 0.011\*federal + 0.011\*leadership + 0.011\*fund + 0.010\*priority + 0.010\*dream + 0.009\*innovation + 0.009\*commitment |

Appendix E

 Permutation test: mean score differences













