
Understanding Presidential Speeches and Executive Orders with Natural Language Processing

Computational Content Analysis

Lily Grier

`lilygrier@uchicago.edu`
The University of Chicago

Linh Dinh

`ldinh@uchicago.edu`
The University of Chicago

Roberto Barroso-Luque

`barrosoluquer@uchicago.edu`
The University of Chicago

Abstract

BLABLABLA

1 Introduction

BLABLABLA

2 Methodology

2.1 Data

We used 2 corpuses for this project: corpus of presidential speeches and corpus of executive orders. The presidential corpus includes 382 presidential speeches from 1963 to 2021 from 12 presidents: Joe Biden, Donald Trump, Barack Obama, George W. Bush, Bill Clinton, George H. W. Bush, Ronald Reagan, Jimmy Carter, Gerald Ford, Richard M. Nixon, Lyndon B. Johnson, and John F. Kennedy. We downloaded this corpus from <https://millercenter.org/the-presidency/presidential-speeches> using scraping methods introduced in week 1. The executive orders corpus includes 1074 executive orders from 1994 to 2021 from five consecutive presidents: Joe Biden, Donald Trump, Barack Obama, George W. Bush, Bill Clinton. We downloaded this corpus from <https://www.federalregister.gov/presidential-documents/executive-orders> using scraping methods introduced in week 1.

2.2 Word Counts and Dependency Parsing

In order to understand the lexical idiosyncrasies and differences between presidential speeches we analyzed individual speeches through finding the most often used words, applying dependency parsing and scoring individual speeches for sentiment. In the coarsest of our analysis we first normalized raw text by tokenizing and stemming individual documents, removing stop words and then counting the number of occurrences for each unique word in each speech. We visualized this analysis by plotting the top ten words used by every president from Lyndon B. Johnson to Joe Biden. Furthermore, we made use of dependency parsing, a text mining technique to capture linguistic dependencies from text. Such methods tag different words in a sentence based on their relationships to other words and the grammatical structure of individual sentences. Dependency parsing gives insights into the dynamic relationship between the different parts of a sentence while also allowing the researcher to understand the complexity of sentences. By using a recursive algorithm, we then navigated all sentences by interpreting them as dependency trees. This recursive analysis allowed us to get the maximum depth of individual sentences, in this paper we interpret dependency depth as language complexity as sentences with more node-levels tend to be syntactically more complex. We then analyzed the minimum, average and maximum sentence depth for all presidents to understand their syntactic style in speeches.

2.3 Sentiment

We made use of sentiment analysis, a natural language processing technique that allows for the systematic identification of affective valence across language to quantify the relative valence of individual speeches across and between presidents. In addition, we made use of sentiment analysis, in conjunction with word embeddings, to understand the relative direction of emotion shown by presidents towards specific policy and legislative problems. It is important to note that in our analysis, as is common in the natural language processing literature, sentiment scores can range between -1 to 1 with negative values denoting detrimental sentiment while positive values signal optimistic sentiment.

2.4 K-Means Clustering

The raw text from each document (e.g., a presidential speech, an executive order) was vectorized using sklearn TfidfVectorizer. Afterwards, the standard K-Means algorithm was applied to the vectorized texts of all documents. The default sklearn hyperparameters other than the number of clusters were used. And, we used grid search with number of clusters 2,3,4 for each of three specifications of data: presidential speeches and executive orders combined, presidential speeches only, and executive orders only. The clustering results (number of clusters) with the most reasonable Silhouette Analysis and Principal Component Analysis (2 components visualized) results were reported in the Results section below. Additionally, for the experiment where we combined presidential speeches and executive

orders, we know in advance the “true” number of clusters; thus, we also evaluated and reported on a few K-Means accuracy metrics such as Homogeneity, Completeness, V-measure, and Adjusted Rand Score.

2.5 Topic Modelling and Latent Dirichlet Allocation

To explore how words clustered together in semantic space, we performed topic modeling on both the executive orders and presidential speeches corpuses. After some preliminary cleaning, we tokenized and normalized our text and applied Latent Dirichlet Allocation (LDA). LDA is a machine learning method that attempts to extract meaningful topics from a set of documents based on the distribution of words within those documents. LDA starts off by randomly assigning each word in a document to a topic according to a pre-specified number of topics. It then uses the probability distribution of topics across each document as well as the probability distribution of words across topics to iteratively change the topics assigned to each word. The algorithm stops re-assigning topics once it reaches a balance of. Because k-means clustering suggested executive orders and presidential speeches are semantically distinct, which aligns with the more technical nature of executive orders compared to the persuasive aims of presidential speeches, we chose to analyze these two categories of documents as two separate corpuses. Preliminary exploration with the executive orders corpus raised the concern that procedural words were creating non-meaningful topics. For example, we had a topic consisting of “shall,” “agency,” “hereof,” and “presidential,” which do not reveal anything about the policy domain in which an executive order lies. To guard against this, we filtered out words common to executive orders that may not be filtered out by the tf-idf cutoff determined by the gensim model. These words included “department,” “shall,” “order,” “federal,” “law,” “president,” “act,” “secretary,” “amendment,” “amend,” “state,” “executive,” “agency,” and “section.” This preliminary step allowed for marginally more meaningful topics found through LDA, though there is certainly room for improvement in effectively filtering out non-meaningful words. For each of our two corpuses, we performed a grid search to find the optimal LDA model specification. As LDA requires the number of topics to be specified in advance, we tried models with 4, 6, 8, 10, 15, and 20 topics. We also varied the range of alpha and beta, trying 0.6, 0.8, and 1.0 for each. Alpha is a hyperparameter which represents topic density. Higher alpha values mean documents will be composed of a greater number of topics. Beta controls the topic-word density. An LDA model with a higher value of beta will produce topics with more similar words to each other. We ranked models based on their coherence scores and performed further analysis on the best-performing model specification from each corpus. The results of the grid searches for each corpus are depicted in Table 1.

Alpha	Beta	num-topics	coherence-score
.6	.6	10	.432312
1	.8	10	.37757
asymmetric	.4	10	.37757
asymmetric	.8	15	.37568
.2	.6	15	.37114
...
.2	symmetric	2	.2598
.4	1	2	.25625
.6	.2	2	.2538

Table 1: Optimization of LDA Hyper-Parameters through Grid Search

2.6 Vector Space Word Embeddings Over Time

The raw text from each document (e.g., a presidential speech, an executive order) was word tokenized and normalized using the word-tokenize and sent-tokenize methods of the lucem-illud package. The sentence structure of each text was retained because it is needed for Gensim Word2Vec’s “continuous bag of words” method for word embeddings. Each unique word in the corpus vocabulary was

vectorized using Word2Vec’s “continuous bag of words” - a neural network model on sequences of texts. In order to identify the semantic changes overtime of focal words relevant to our project, which are “president,” “regulation,” “economy,” “health,” “energy,” “equity,” we needed to analyze the word vector embeddings of these focal words in their semantic space over time. Thus, we first generated one word vector embeddings model for each year using all the tokenized corpus (of both presidential speeches and executive orders) in that year. We chose to combine both presidential speeches and executive orders into 1 single corpus for this analysis because our hypothesis was that the semantic variations can occur in both corpus. Additionally, combining corpuses also allowed us a wider vocabulary of common words (of all years) to investigate. Given that our corpuses were small in size, we realized that a lot of the words (topics) of our interests, which admittedly are more recent “hot” topics, were not present in one or both of our corpuses until very recent years. Afterwards, we aligned the dimensions of the word vector embeddings for all words in the vocabulary that are common in all years. Finally, for each focal word, we calculated the cosine similarities of the vector embeddings between years and visualized them (more specifically, we visualized 1 - cosine similarity) in a heatmap. For this analysis, we restricted the time period to the last 20 years (2000-2020) due to the fact that some of our focal words did not appear until more recent years.

2.7 Doc2Vec

We employed doc2vec on the executive orders corpus to further explore the relationships between documents. Doc2vec is a method that creates numeric representations of documents. We created the following tags as keywords for our documents: “energy,” “environment,” “education,” “labor,” “employment,” “health,” “healthcare,” “defense,” “security,” and “war.” We tagged documents with the keywords they contained, then fed the tagged orders into gensim’s doc2vec model. We also used doc2vec to calculate the cosine similarity between documents and to find documents similar to certain words (e.g., “protect,” “sustainable,” “medicare,” “tax”) as calculated by their vector embeddings.

3 Results

3.1 Word Counts, Dependency Parsing and Sentiment

To gain a coarse understanding of presidential rhetoric style we counted the number of unique words by each president (Fig 1a). In order to give context, we also counted the number of speeches by each president present in our corpus. In general it seems that the number of unique words is mostly a function of the number of speeches given by a president. For example while Joseph R. Biden is by far the president with the lowest number of unique words used, he is also the president with the smallest number of speeches present in our corpus (just 1). Conversely, Lyndon B Johnson who used more than 11,000 unique words across his speeches did so throughout 70 speeches. Nonetheless, it is worth noting that both Reagan and Obama, often the most celebrated presidents of their respective parties, made use of more than 12,000 and 11,000 unique words, respectively, during many less speeches than Johnson. Our word count analysis, surprisingly, showed that despite party affiliation or era, most presidents seem to use a similar set of ten top words by raw count (Fig 1b). Staple speech words such as america, nation, democracy, united, etc are used frequently by all presidents of both parties. Such regularity, however, does not hold for the sentence complexity across presidents speeches (Fig 1c). While the average sentence depth of most presidents ranges between 2 and 5 levels, maximum depth has high variance. In the lower end of the spectrum, presidents such as Donald Trump and George W. Bush makes limited use of complicated sentences reaching a maximum of 12 or less in dependency depth. In contrast, president Obama, Clinton, George H.W. Bush and Ford reach a maximum dependency depth of more than 16 levels each. Such differences in sentence complexity illustrates the idiosyncratic style of individual presidents. In terms of relative valence in presidential speeches (Fig 1d) we see that, in general, presidents tend to be upbeat in their speeches with a vast majority of speeches in our corpus scoring the maximum possible value of 1. Notably, it was republican presidents Donald Trump and George W. Bush who struck a pessimistic tone more often than any other president with each having more than 5 speeches with negative sentiment scores. This result is consistent with the political punditry accounts of both presidents presenting a more cynical and fatalistic world outlook than their counterparts.

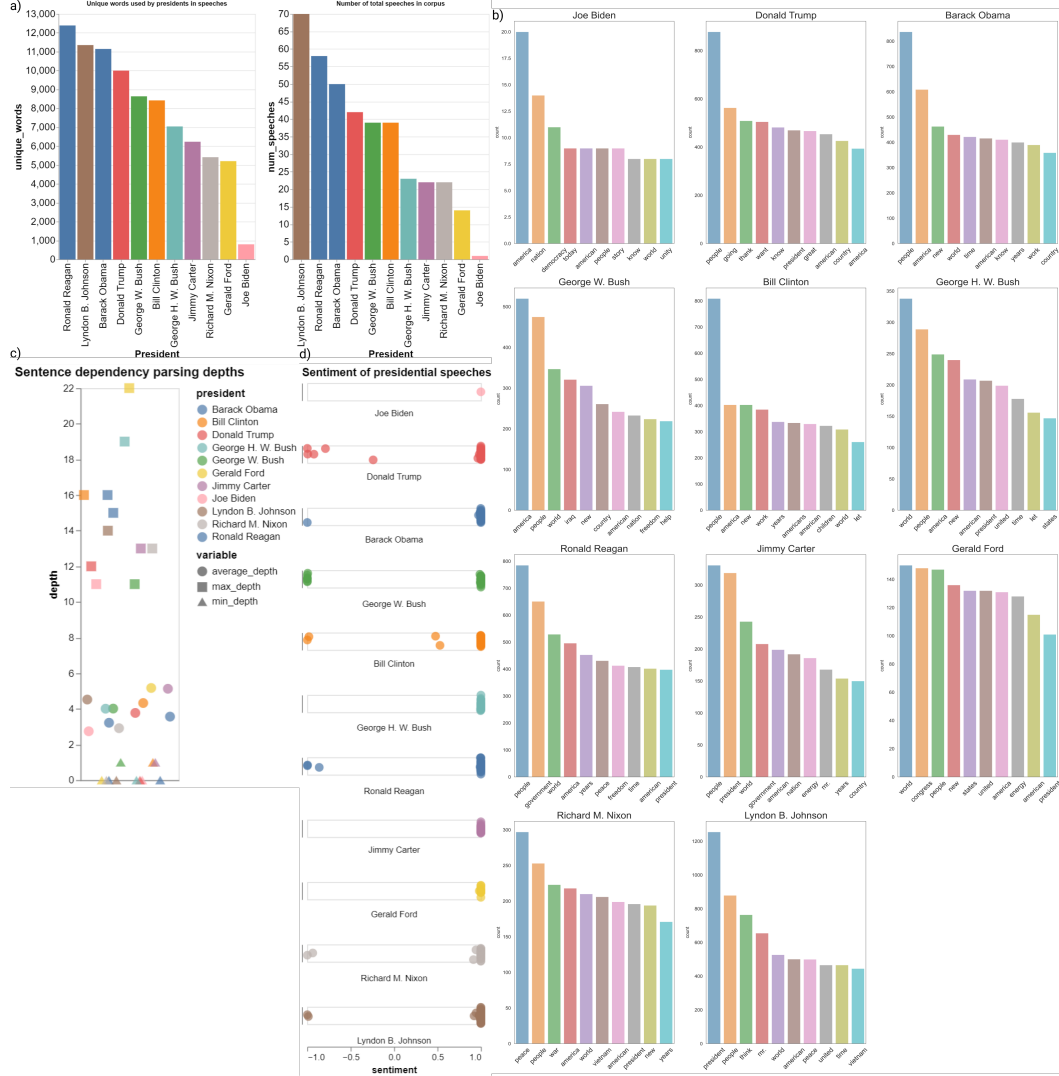


Figure 1: Analysis of Presidential Speeches

3.2 K-Means Clustering

For the data that combined presidential speeches and executive orders, based on Silhouette Analysis, the best number of clusters was 2, which also correctly distinguished between presidential speeches and executive orders as seen in by plotting the 2-first principal components (Figure 2, top left and right figures). Since we have the “true” clusters in this experiment, we also looked at a few K-Means accuracy metrics, which were all very high and indicated that the K-Means algorithm worked very well with this classification task (i.e., identifying presidential speeches vs. executive orders). The Homogeneity is 0.962, Completeness is 0.968, V-measure is 0.965, and Adjusted Rand Score is 0.986. This result also suggested that the semantic characteristics of presidential speeches are very different from that of executive orders and vice versa. For the data that contained only presidential speeches or only executive orders, based on Silhouette Analysis, the best number of clusters were also 2 for both datasets. The Principal Component Analysis (2 components) plots also looked reasonable (the bottom left and right figures). We do not have the “true” clusters in these 2 experiments; thus, we could not look at any K-Means accuracy metrics as before. However, the semantic separations within each corpus (i.e., presidential speeches alone and executive orders alone) were investigated with topic modeling and the results were presented in the following sections.

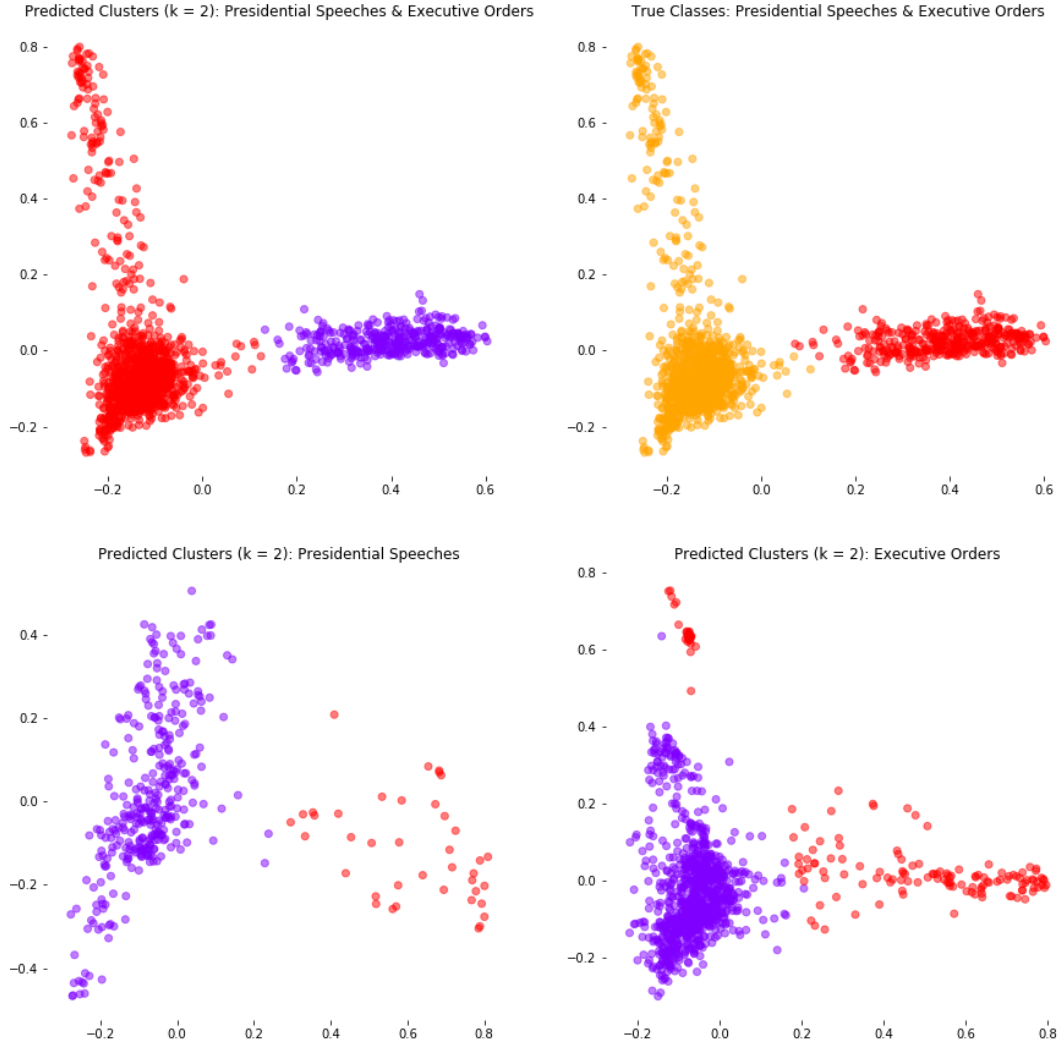


Figure 2: K-Means through PCA, Speeches and Executive Orders

3.3 Topic Modeling and Latent Dirichlet Allocation

Presidential Speeches LDA Our optimal LDA model for the presidential speeches corpus, with a coherence score of 0.432, had 10 topics, an alpha value of 0.6 and a beta value of 0.6. The top 10 words for each cluster are shown in supplementary figure 2. We assigned each topic an informal label based on its most prominent words and our knowledge of common topics within presidential speeches in Table 2. When we looked at the breakdown of topics across individual speeches, we saw that each president in our corpus mainly spoke about topic 1 (welfare and aid) and topic 8 (diversity and inclusion) in his first speech as president, which is presumably the inaugural address. This makes sense, as these topics are about helping those in need and embracing the American public and can be framed in party-neutral ways. Inaugural speeches generally focus on abstract American ideals and less on concrete policies, talking about broad topics such as peace and the strength of the American people rather than war or health insurance. Visualizing the breakdown of topics by president provides insight into variations in policy focuses. Below (Figure 3 part a) we have topic distributions for Barack Obama and George W. Bush. Obama's speeches cluster around topic 9, the people's president, which Bush's speeches center around topic 2, America and the world, which makes sense as Bush's presidency is characterized by the wars in Iraq and Afghanistan and growing fears around national security following the 9/11 attacks. We also see health insurance as a prominent topic in Obama's

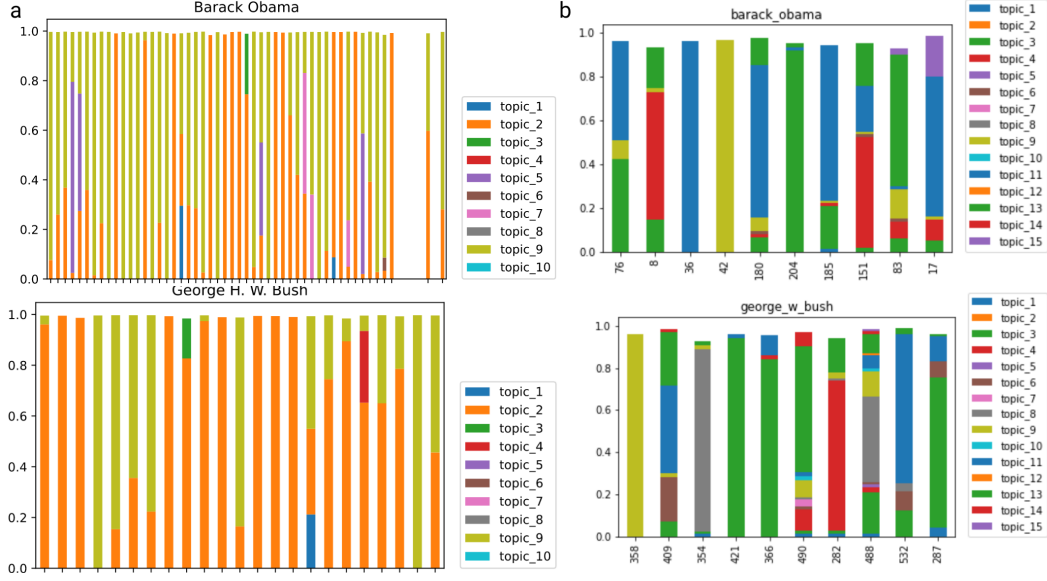


Figure 3: Distribution of topics in: a) presidential speeches b) Executive orders

speeches, which aligns with the narrative of the passing of the Affordable Care Act as Obama’s most notable accomplishment.

Topic id	Topic name
1	Welfare and aid
2	America and the world
3	Foreign wars
4	Great power rivalry
5	Celebrations
6	Supreme Court
7	Health insurance
8	Diversity and inclusion
9	The people’s president
10	NATO and proxy wars

Table 2: Informal Labels for Optimized LDA Speeches Model

Executive Orders LDA The optimal LDA model for the executive orders corpus, with a coherence score of 0.542, had 20 topics, an alpha value of 0.8, and a beta value of 1. The topics with their informal names and top keywords are in Table 3.

Topic id	Topic name	Salient Words
1	Military procedures	Follow, court, martial, military, punishment
2	Health insurance	Health, care medicare, affordable, coverage
3	Payment systems	Schedule, pay, rate, service, adjustment
4	International crime	Turkmenistan, interpol, combatant, vested
5	National security/intelligence	Security, information, national, intelligence, homeland
6	Census	Census, race, population, status, sex, apportionment
7	Crisis	Pursuant, emergency, foreign, prohibit, block
8	Life Sciences/Biology	Genetic, mutation, diagnose, metabolic, chromosomal
9	Monuments	Statue, monument, national, hero, Washington, memorial
10	International crime	Turkmenistan, vested, interpol, combatant
11	Governmental bureaucracy	Committee, register, commission, member, advisory
12	Internet regulation	Platform, online, speech, restrict, deceptive, ftc, twitter
13	Space	Space, command, nuclear, weather, exploration, moon, nasa
14	Public health	Agency, health, public, covid, community, program
15	Foreign diplomacy	Immunity, international, binational, israel, european
16	Labor issues	Board, dispute, railroad, employee, company, union, represent
17	Military honors	Medal, award, arm, navy, guard, coast, terrorism, war, iraq, afghanistan
18	People in authority	Employee, director, general, title, head, officer
19	National security	Information, classify, classification, security, national, provide
20	Energy production	Contractor, energy, facility, use, cost, vehicle, environmental, chemical

Table 3: Informal Labels for Optimized LDA Executive Orders Model

This topic model had some limitations. First, two sets of the topics produced were nearly indistinguishable from one another. Topics 4 and 10 both centered around international crime, while topics 5 and 19 both centered around national security and intelligence. The grid search jumped from 15 to 20 topics, so future research could test topic numbers in between to see if a more optimal topic distribution exists. Further, topics 11 and 18, loosely titled “governmental bureaucracy” and “people in authority,” respectively, are not very informative, as they mostly contain procedural words (e.g., “committee,” “member,” “advisory,” “director,” “officer,” etc.).

Despite its limitations, the topic model does convey some useful insights about our corpus of executive orders. We see major topics of executive orders such as health insurance, scientific research, and foreign diplomacy. Interestingly, healthcare clustered into two distinct topics: one focusing on public health (including COVID-19) and associated infrastructure and the other associated with health insurance. That the model was able to pick up on these contextual nuances speaks to its power in helping us better understand the semantic landscape of executive orders. It makes sense that payments were their own topic, as an order entitled Adjustment of Certain Rates of Pay which modifies the wages of federal employees gets passed almost every single year.

The topics in this model suggest that executive orders focus more on foreign policy, healthcare, and national priorities such as research than on matters such as labor and regulation of businesses. Perhaps because labor laws vary across state jurisdictions and there is not much tangible policy the executive branch can enact on this front, executive orders don’t hone in on domestic labor and economic policy. We can use these topic models to visualize the topics that different presidents emphasize.

We coded each order by its dominant topic, then plotted the distribution of dominant topics for each president. Across all presidents, we see our “filler” topics of governmental bureaucracy and people in authority appearing as frequent dominant topics. Future iterations upon this work should try to

eliminate these topics through tf-idf filtering or deleting non-meaningful topics and then re-classifying documents into the existing topics. Notably, the public health topic and not health insurance was prominent in Obama’s executive orders, though health insurance was a major topic in his speeches. Because the actual passing of the Affordable Care Act occurred through congressional legislation and not through executive orders, it is probable that Obama’s orders made more general declarations about public health than they supported specific insurance policies or healthcare systems. For George W. Bush, we see crisis, national security, and public health emerge as major topics. Bush served as president during major crises such as Hurricane Katrina and the 9/11 attacks and did indeed have many orders aimed toward various mechanisms for homeland security. Though the dataset contains only the first two months of Biden’s presidency, we see the public health topic dominating most of his 37 orders, which fits for a president who came into office during the height of a pandemic.

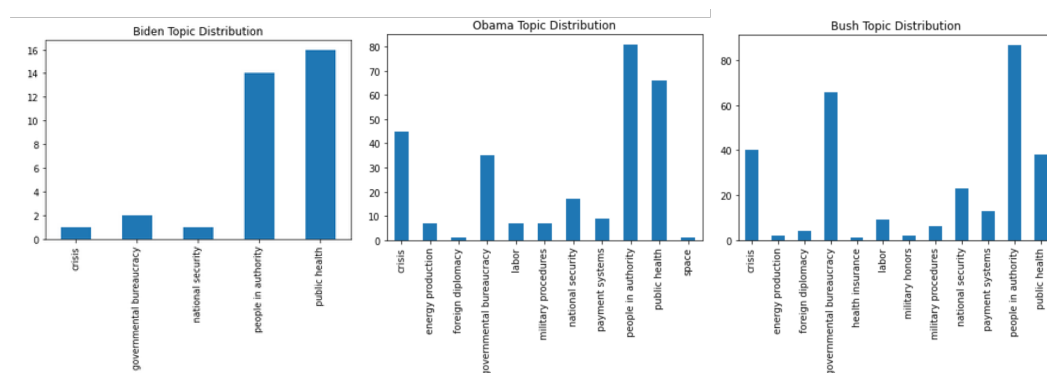


Figure 4: Executive orders LDA Topics distribution

3.4 Vector Space Word Embeddings Over Time

Interestingly, as mentioned above, some of the words (topics) of our interests such as “climate” or “environment” were not present in the yearly corpus of presidential speeches and executive orders until very recent years (i.e., after 2010). Hence, another set of analyses into the appearance and disappearance of words in these 2 corpuses could be interesting; however, the scope of this paper does not address this issue. As seen from the above figures, a lot of the semantic meaning of our focal words, including “president”, “regulation”, “economy”, and “equity”, had a significant change in 2017, which coincided with the inauguration of Donald Trump. This result suggested that, as expected, Donald Trump is quite different from the other presidents, at least in the usage of words and vocabularies in presidential speeches and executive orders. The significant change of the focal word “equity” in 2017 was particularly interesting. Another interesting observation was that there were a lot of big semantic changes in the year 2020, at least among the focal words we chose to look at, which includes “president”, “regulation”, “economy”, “health”, and “equity”. This result seems to suggest that 2020 was indeed a “special” year. We anticipated that the events and sentiments surrounding Covid-19 probably impacted these semantic evolvments. This hypothesis was further corroborated via the fact that the word “health” had a sudden significant change in its semantic embedding in 2020. Before 2020, the word “health” had a relatively stable semantic embedding throughout the past 20 years. Ideally, additional investigations into the potential drivers of these semantic changes could be interesting (such as just simple word counts in case words such as “health” and “equity” appear more often in 2020). One last observation was how the word “economy” started to fluctuate in its semantic space, which was slightly before the 2008 financial recession. This, interestingly, suggested that investigating the word vector embeddings of presidential speeches and executive orders somehow was able to anticipate the financial crisis in 2008. Thus, further analyses into this space could also be quite interesting.

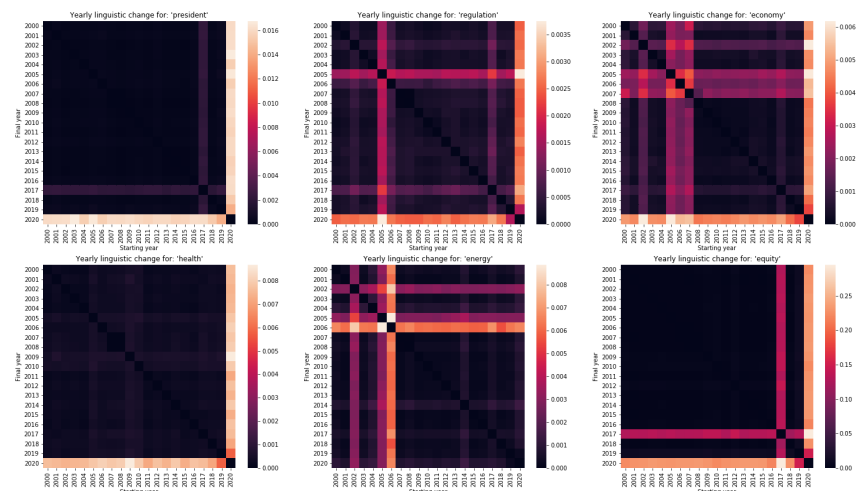


Figure 5: Word Embeddings, visualizing word change use over time

3.5 Doc2Vec

As expected, all orders in our executive order corpus exhibited relatively high cosine similarity, and a simple document-document comparison failed to identify meaningful differences between orders. Because executive orders tend to be largely similar regardless of their content focus and follow a rigid structure, doc2vec may not pick up more nuanced similarities. Attempting to parse out the content of executive orders from their overall structure could be an avenue for future research. We also used doc2vec to find the most similar words to certain document embeddings. Combating Race and Sex Stereotyping, an infamous order enacted by Donald Trump that argued against diversity, equity, and inclusion initiatives that talk about structural racism, exhibited similarity to “contractor,” “employee,” “employment,” “position,” and “appointee.” Interestingly, the main words in this topic don’t have to do with race or gender but rather the workplaces in which these trainings take place. This may be because a greater number of executive orders talk about employees than diversity initiatives, so the model picked up on these procedural elements more. Obama’s 2016 order, Establishing a Community Solutions Council, mapped onto “attend,” “representation,” “partnership,” “diverse,” and “patriotic.” It called for a council to oversee and support community-based initiatives ranging from workforce training programs to food access to environmental justice. The word mappings here are not surprising, as they refer mostly to the broader goals and motivations of the act rather than the wide-ranging policy areas. As a third illustrative example of document-to-word mappings, we looked at Biden’s 2021 act entitled Rebuilding and Enhancing Programs to Resettle Refugees and Planning for the Impact of Climate Change Migration. The act mapped onto “usrap” (an acronym for United States Refugee Admissions Program), “deficiency,” “aggressive,” “refugee,” and “crucial.” The words “deficiency,” “aggressive,” and “crucial” do not actually appear in the text of this order, so it is notable that they came up as having similar embeddings. The order emphasizes the need for the United States to increase its capacity for taking in refugees and the heightened urgency of this need given climate change displacing many individuals around the world. Thus, it makes sense that an emphasis on inadequate capacity would link to “deficiency” and stressing the need to enhance our efforts in a timely manner would link to “aggressive” and “crucial.” The figure below depicts 10 orders selected at random and shows their cosine similarity to each of our 10 keywords. Some of these mappings are more accurate than others. For example, Promoting Beautiful Federal Civic Architecture exhibits puzzling similarity with war and environment. The order establishing a board to investigate labor disputes with the Southeastern Pennsylvania Transportation Authority maps onto labor and employment as expected, but also exhibits strong similarity to security. Thus, there is reason to doubt these mappings are robust enough to classify unlabeled documents. This lack of robustness

could be due to executive orders having similar forms and focusing more on specifying administrative procedures rather than on domain-specific policy substance. Future research should seek to better understand the quality these embeddings, especially for topics such as education and labor, which are less commonly addressed by executive orders. We also used our doc2vec model to find orders that mapped most strongly onto certain words. For example, we saw that “sustainable” mapped onto orders establishing the Great Lakes Interagency Task Force, managing forests and federal lands, and supporting the use of space resources. Similarly, “tax” mapped onto orders promoting the purchasing of American-made products, encouraging the use of energy efficient devices, making changes to regulation, and supporting the growth of the economy. Tax also exhibited strong similarity to an order restricting access to abortions, and upon inspection of the order’s text, we found that the order prohibits the use of tax credits to pay for abortions under the Affordable Care Act. Doc2vec also proves to be useful for identifying executive orders that map onto combinations of keyword embeddings. For example, adding together the vector embeddings for “land,” “protect,” and “water” fetched orders on protecting and managing forests and trails, reducing wildfire risk, restoring the gulf coast, and recovering space resources. Though this preliminary analysis is not exhaustive, it gives us reason to believe doc2vec could be an effective tool to identify orders that map onto keywords.



Figure 6: Cosine Similarity in Executive Orders

4 Conclusion, further research and limitations

Our work seeks to make inferences about relationships between the Democratic and Republican parties in the United States, the role they play in a democratic society, and the impact presidential speeches and executive orders have on policy changes. Our analysis is limited in that we are only including presidents in our corpus; future work could include other elected officials. It would be interesting to look at House members who draft legislation and the ways they communicate about policy issues (or how Senators vote on issues and the speeches they make to their home states). Our work will reveal the issues that were important to the American public to the extent that those issues are reflected in presidential rhetoric and action, but adding on a layer of public sentiment (e.g., polling results, news articles, social media) in future research would better contextualize our findings. Future investigations could also include presidential communications such as press releases and social media activity (i.e., Twitter) to develop a more thorough understanding of the relationship between presidential rhetoric and legislation.

blablabla.....

5 References

- [1] Chipman H. & Gu H. (2006): Interpretable dimension reduction. *Journal of Applied Statistics*
- [2] Ding J. , Condon A. & P.Shah S. (2018). Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nature Communications*
- [3] Hosseini B. & Hammer B (2019): Interpretable Discriminative Dimensionality Reduction and Feature Selection on the Manifold. *BiorXiv*

6 Supplementary Figures

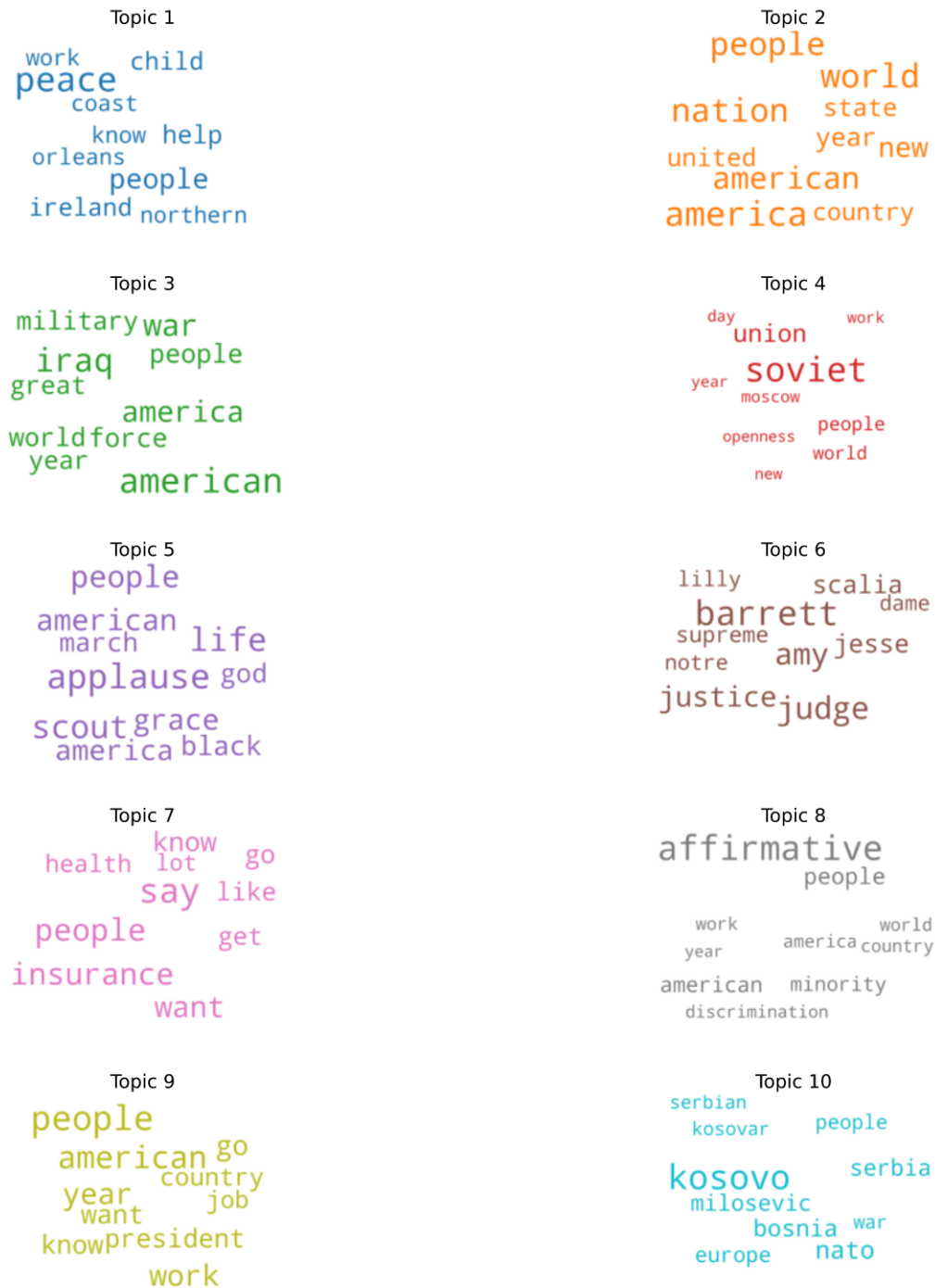


Figure 7: Word Clouds of LDA Topic Models for presidential speeches



Figure 8: Word Clouds of LDA Topic Models for executive orders