Underpinnings of Political Leaning: Using Collocation Extraction and Semantic Analysis to Categorize Ideological Frames

Kavi Mehta

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2019-89
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-89.html

May 20, 2019

Acknowledgement

Thank you to Professor Jerome Feldman and Professor Dekai Wu for their guidance and support throughout this project.

Underpinnings of Political Leaning: Using Collocation Extraction and Semantic Analysis to Categorize Ideological Frames

Kavi Mehta

Research Project

Submitted to the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, in partial satisfaction of the requirements for the degree of **Master of Science, Plan II**.

Approval for the Report and Comprehensive Examination:

Committee

Jerome A Feldman

Professor Jerome Feldman
Research Advisor

5/17/2019

(Date)

* * * * * * *

Professor Dekai Wu
Second Reader

May 17, 2019

(Date)

17 May 2019

The language used in political discourse varies depending on the worldview of the speaker, and we describe specific instances of politically charged phrases as political frames. A frame is a group of words that appears consistently across political corpora, and is applied to evoke particular emotions or opinions regarding an issue. Such frames get used in all language, whether it is intentional or not. We propose a system that can automatically pull out frames from a corpus using statistical and linguistic analysis. The result of this system is a tool that ingests large amounts of texts and constructs frame matrices — an organizational structure to categorize and document political frames. Overall, the main contribution is in demonstrating the value of such a system and outlining the means through which it can be done.

Contents

1 Introduction **6**

2 Preliminary Exploration **8**

2.1 Liberal - Conservative Classification Model . . . 8

2.1.1 Methods . . . 8

2.1.2 Results . . . 9

2.1.3 Discussion . . . 15

2.2 Manual Frame Matrix Construction . . . 16

2.2.1 Methods . . . 16

2.2.2 Results . . . 17

2.2.3 Discussion . . . 17

3 Project Goals and Outline **21**

3.1 Goals . . . 21

3.2 Outline . . . 21

4 Project Work **23**

4.1 Data Collection . . . 23

4.2 NP/SF Sorting . . . 23

4.3 Collocation Extraction . . . 24

4.3.1 Definition . . . 24

4.3.2 Algorithm . . . 24

4.3.3 Determining Interesting Words . . . 27

4.3.4 Collocation Results . . . 27

4.4 Collocation Filtering . . . 27

4.5 Key Sentence Collection . . . 29

4.6 Frame Analysis . . . 29

4.7 Output Synthesis . . . 32

5 **Analysis and Results** **34**
5.1 Algorithm 34
5.2 Qualitative Analysis 36
5.3 Quantitative Analysis 36
5.3.1 Experimental Setup 37
5.3.2 Experiment Results 37
5.3.3 Resulting Matrix 39

6 **Discussion** **40**

7 **Further Work** **42**

Bibliography **43**

1 Introduction

The discourse surrounding political topics and contemporary issues changes dramatically depending on the speaker. Conversations take on different forms in different parts of the country, where information is disseminated through entirely disparate news sources and prominent individuals. Depending on who one engages with, the questions surrounding immigration, climate change, civil rights, and more come with firm undergirding loaded with inconsistent context and assumptions. In short, the language around politics is multifaceted and highly dynamic, the shape and style altered in every community.

For many, this phenomenon is referred to as "bias". In cognitive science, bias refers to a deviation from the norm [12], and with regards to news and information, people take this to mean a deviation from their perception of the raw facts. How information gets skewed or altered greatly affects how people feel about and understand the issues at hand [25], so the claim that news bias has become pervasive in modern media should not be taken lightly. Furthermore, psychologists describe how sources fall victim to selection bias, which affects how people choose what information to share, and presentation bias, which affects how people present the knowledge they have chosen [8]. All combined, these effects create the perception that news sources are inherently "biased", and that their discussion of topics is simply a manifestation of these factors, not a reflection of how people think about these issues [17].

But the issue of "bias" — or, more accurately, the difference in how issues are discussed across news sources — runs deeper than just how information is presented; the way people talk about issues is dependent on their worldview and values [1]. This idea, which is relatively new in the fields of cognitive science and linguistics, has been best captured by Professor George Lakoff in his book *Don't Think of an Elephant: Know Your Values and Frame the Debate* [11]. Lakoff describes the concept of political framing, which is the use of particular language to evoke certain thoughts and emotions in the audience. He argues that people and the news do this constantly, whether consciously or not. Consider, for example, how the common conservative phrase "tax burden" associates taxes with something negative, reinforcing the notion that taxes are inherently bad and should be reduced. In using this language, people reflect how they feel about taxes, and writers in the news use such language to better connect to their target audience [6] [19].

Lakoff explains how there are two differing fundamental worldviews which inform people's opinions and attitudes towards political information. He dubs these categories Strict Father and Nurturant Parent. The names are derived from the idea that people's feelings largely come from their family structure, and generally one can distinguish the two as such: Strict Father entails the more hardline attitudes and notion that one should be able to support themselves, whereas Nurturant Parent encourages a broader sense of forgiveness and help. Largely, these categories align with America's current political factions, with Strict Father mindsets leaning conservative and Nurturant Parent mindsets leaning liberal.

In this light, we can shift our focus away from the abstract notion of bias and towards the discovery of different political frames — specifically, identifying frames from either a Strict Father or Nurturant Parent mentality. In uncovering what frames different people use to discuss different issues, we can identify what language people use when thinking about various political topics. In this project, we set out to create a system that can ingest corpora and use a combination of statistical analysis, machine learning, and natural language understanding to pull out political frames on different issues for each of the framing worldviews.

2 Preliminary Exploration

2.1 Liberal - Conservative Classification Model

To begin understanding how a worldview and political leaning manifest in news media, we initially sought to create a discriminator between liberal and conservative news sources. Such a classifier has uses in determining the partisanship of any political corpus, regardless of its origin. This is valuable because while it is easy and well-documented to observe the inclinations of large publications — for example, how Breitbart leans heavily conservative and the Huffington Post leans heavily liberal — it is not as clear how one can determine the leaning of smaller publications or even individual posts and speeches. Along these lines, we hypothesized that it is possible to create a classifier to determine the news source of an article given its contents based on the word frequency.

2.1.1 Methods

The methodology for this project falls into three sections: the data collection, the word frequency analysis, and the model training.

Data Collection

Unfortunately, there does not exist a publicly available large dataset of political articles, so we had to explore alternative methods of curating a dataset of our own. Ultimately, we settled upon using the eventregistry.org API, an archive of articles that updates its database every day, allowing us to get current data. These experiments were conducted in the Fall of 2017.

For news sources, we wanted to choose a relatively small set with an even distribution of conservative and progressive biases. Using our own knowledge about popular news sources and their reputations, we selected the following news sources: Breitbart News, Fox News, The National Review, The New York Times, San Francisco Gate, and The Washington Post. In total, we curated roughly 1800 articles.

Word Frequency

Rather than grouping all of the words together, we chose to analyze the article data in two separate paths, one using article titles and one using article bodies. We presumed that titles capture the important information of the article and summarize its point, while the bodies dive deep into the specifics and are more likely to be flooded with extraneous words. Furthermore, grouping the bodies and titles together would cause the title data to be drowned out by the magnitude of the body text, and would therefore lose valuable information about which words were selected to be in the title.

We counted the occurrence of words in each article and in each source, and removed stop words (words like "the" and "a" that trivially appear with great frequency but do not yield any insight) from the data. Across the sources, we compared the most common title words and body words.

Additionally, for each word, we computed the variance of that word's frequency across the sources. To do this, we normalized the count of each word for each source using the number of articles for that source, and calculated the variance of the resulting values.

Model Training

In order to create features for the article titles and bodies, we utilized the word count data from the previous section. For the titles, we found the 3000 most common words across all of the titles and created feature vectors that describe the number of of occurrences of each of the most common words in a given title instance. For example, the title "President Trump" would have a length 3000 feature vector consisting entirely of 0's except for a 1 indicating a single occurrence of the word "president" and another 1 indicating a single occurrence of the word "trump." The same procedure was used for article bodies.

We chose to train two types of models for both the titles and the bodies: a Decision Tree classifier and a Gaussian Naive Bayes classifier. We measured the accuracy of the models when they were trained using only subsets of the features—from 1 feature (the single most common word) all the way up to the complete 3000.

2.1.2 Results

The following plots show the results of each of the described analyses: most common title words by source, most common body words by source, the highest variance words for both the titles and bodies, and the accuracy of the different models in classifying sources based on title or body features. Note: While some of the plots reveal a difference in the number of articles curated for each source, these variations were normalized for later analyses as described in the methods.

Most Common Words in Article Titles

In the article titles, we observe a number of trends. First of all, almost all of the sources have "trump" as the most common word in titles—for obvious reasons, this is understandable and expected. In the National Review articles, however, "trump" is supplanted by a number of other words, and this suggests that the National Review tends to focus on other topics.

Another interesting finding is that all of the sources have a very similar set of common words, indicating that on average, the words in article titles do not differ by all that much. While there are differences, we overall notice the same topics and themes (e.g. one source containing "sexual" and another containing "harassment").

Most Common Words in Article Bodies

In the article bodies, we observe similar trends to that of the titles. Namely, almost all of the sources have "trump" as the most common word in bodies.

We also notice that many of the most common words are relatively uninteresting, words like "more" and "up", which do not reveal much about the source's writing or bias. These could be filtered out using a more comprehensive stop word list, but there is a trade off between having too few stop words, which sometimes yields uninteresting words, and having too many stop words, which may oversimplify the data.

Highest Variance Words Across Titles and Bodies

The variance across words is one of the most interesting plots: it reveals which words are the most important in distinguishing sources. In the bodies, for example, we see many words relating to religion and Christianity ("church", "christian", "catholic"). This makes sense—conservative sources tend to discuss these topics frequently, as it is an important subject to their audience, while progressive sources hardly discuss them at all.

Decision Tree and Gaussian Naive Bayes Accuracy

The model accuracy plots highlight multiple noteworthy trends. First, we see that it takes relatively few features before the models' improvement tapers off; the second plot emphasizes this by zooming in on only the first 250 features. Second, we notice that the models only ever achieve 50% accuracy, which is significantly better than random guessing, but overall not very high.

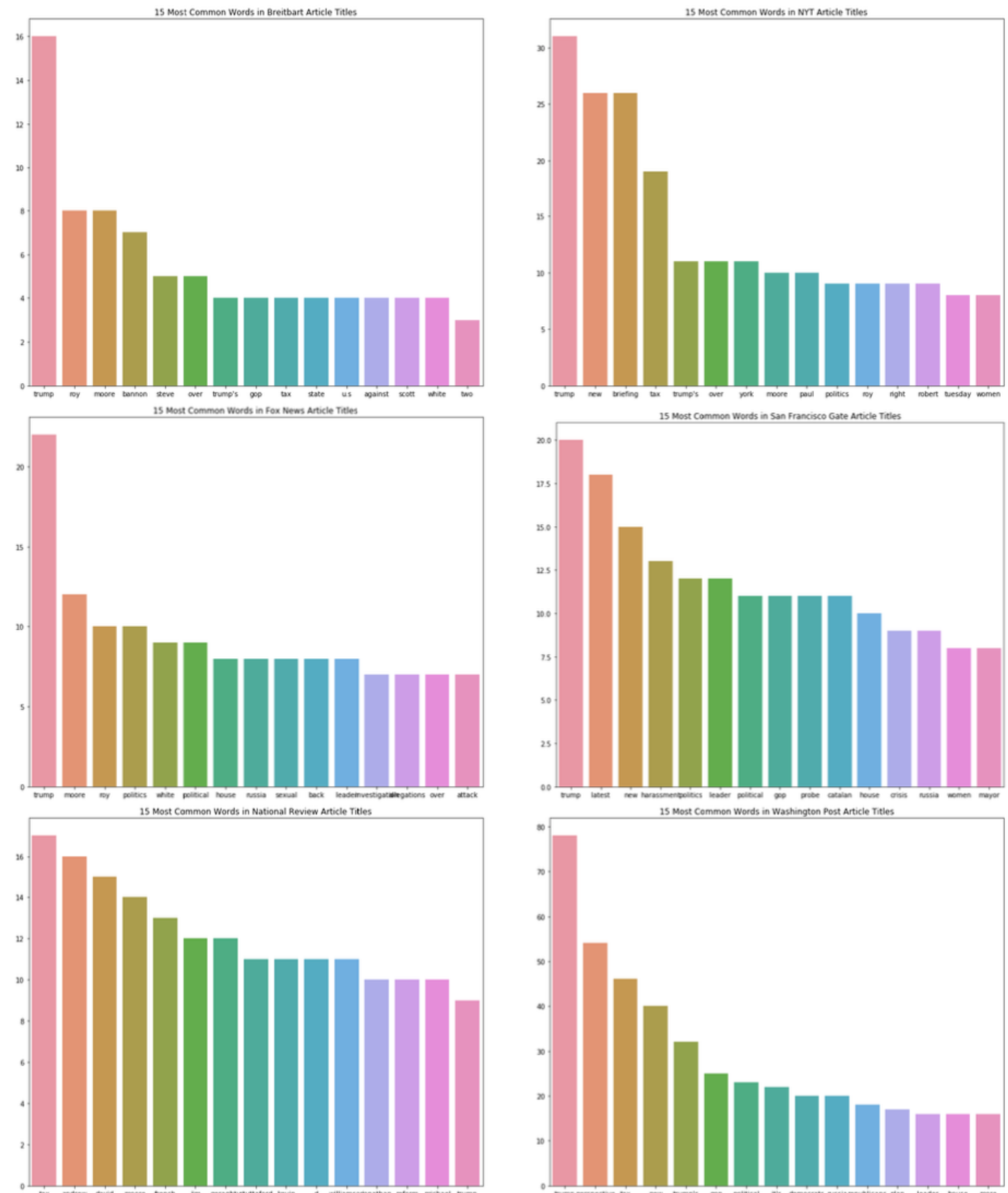


Figure 2.1: Most common words in article bodies

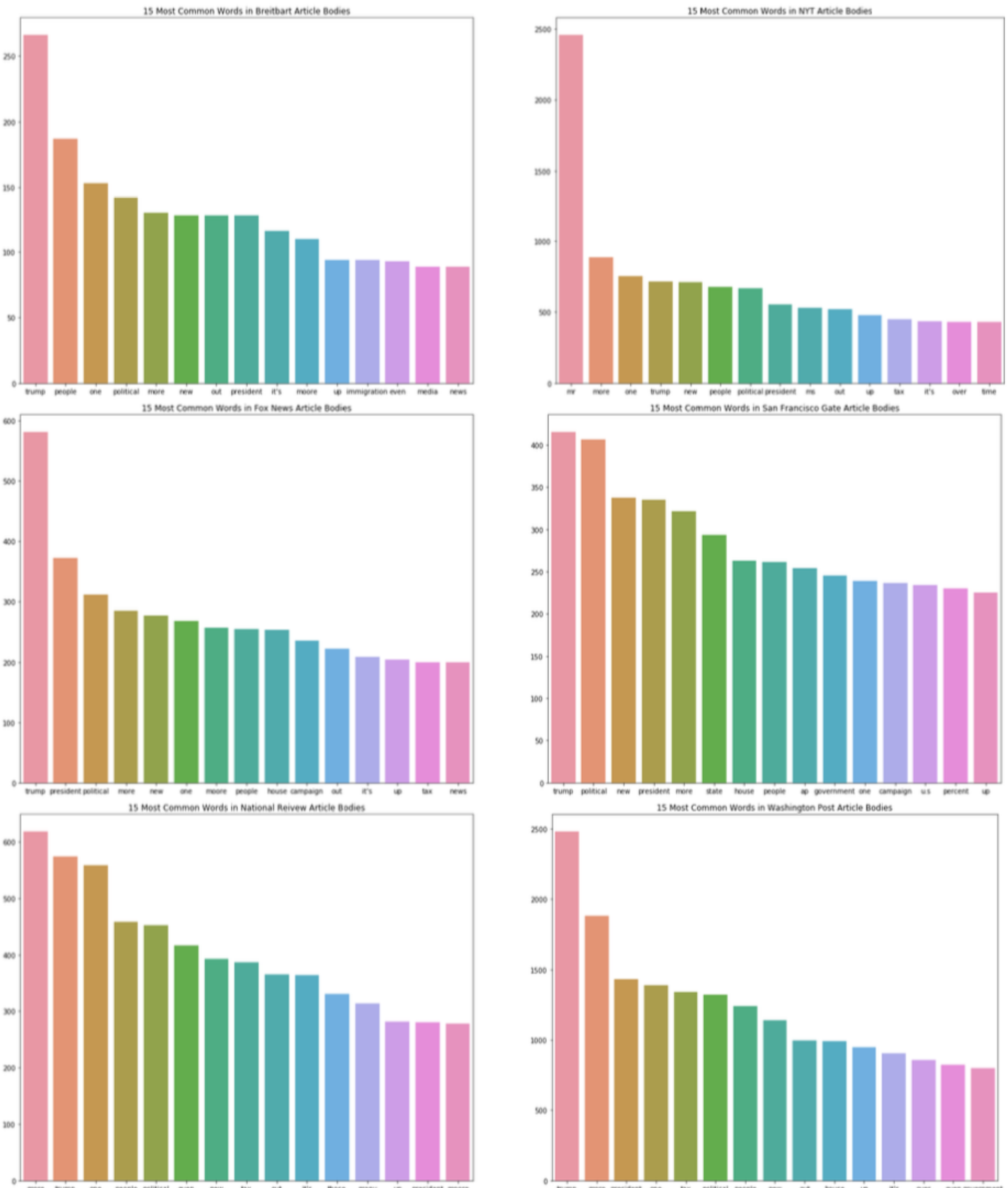


Figure 2.2: Most common words in article titles

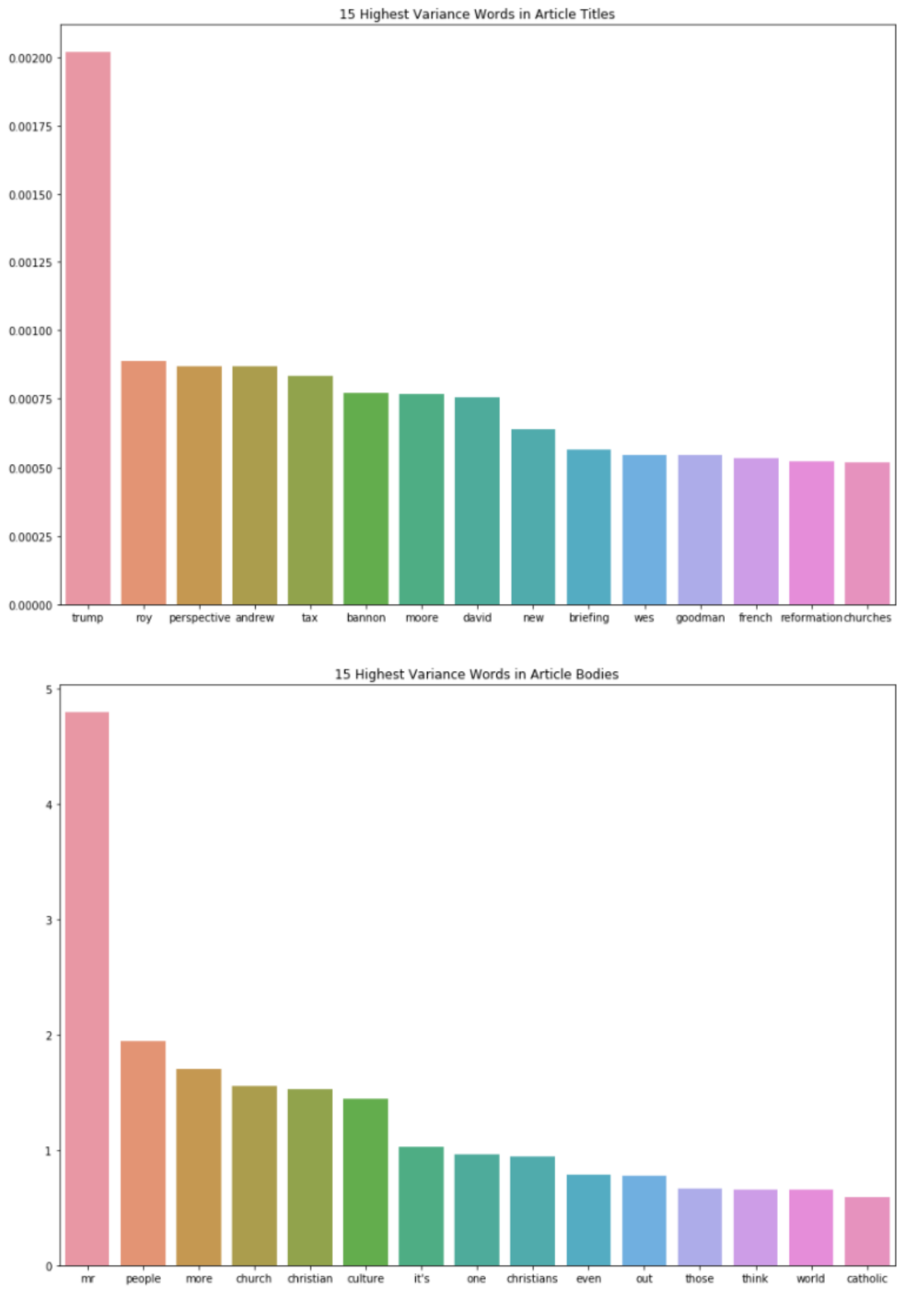
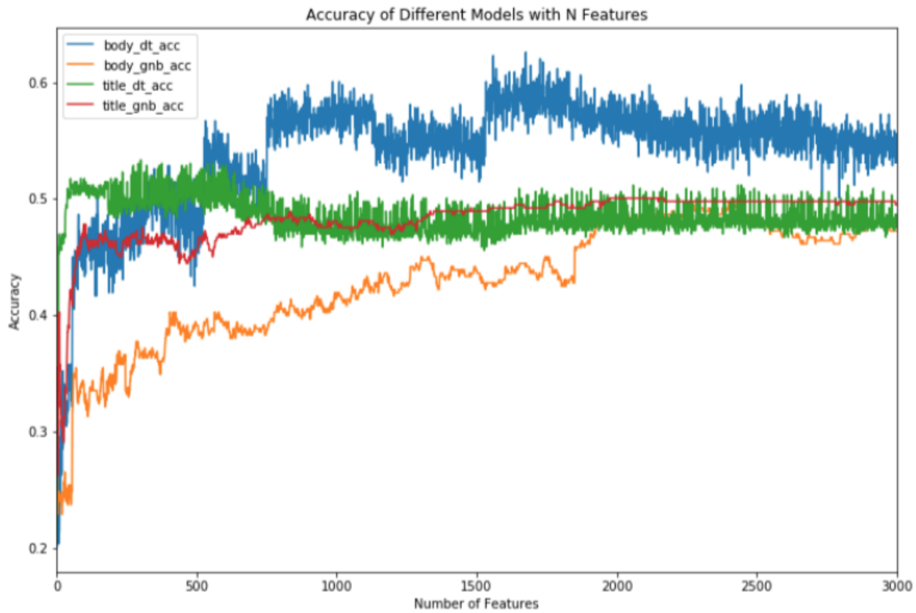


Figure 2.3: Highest variance words



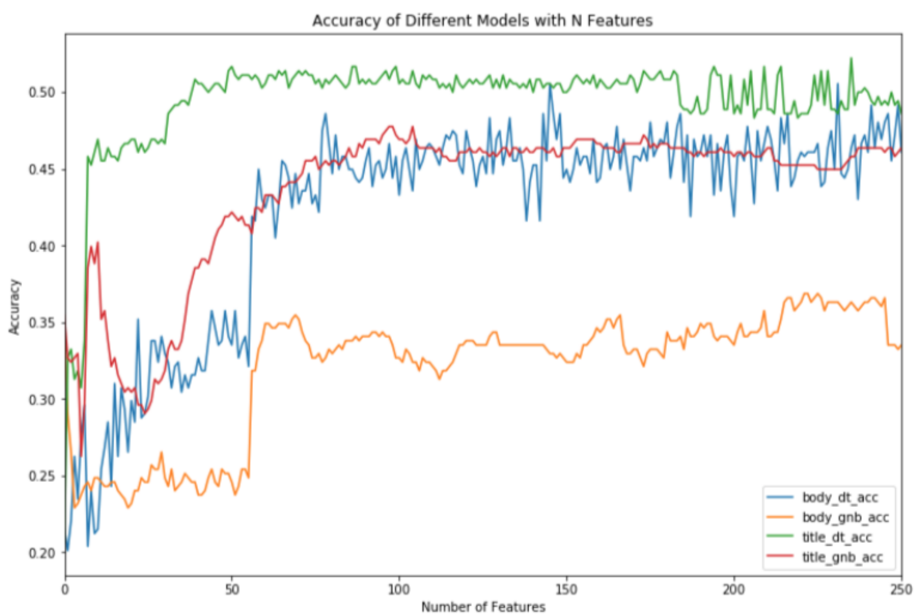


Figure 2.4: Model accuracy

2.1.3 Discussion

The results highlight a number of apparent differences between various news sources, and indicate that sources are indeed biased in order to appeal to their audience and promote particular opinions. Just from the most common words in the article titles and bodies, we clearly see the primary media biases discussed in the introduction: selection bias and presentation bias. As stated, the National Review interestingly does not have "trump" as its most common title word, which shows that they focus on other topics and exhibit bias towards other stories. Comparing the common words in Fox News titles and San Francisco Gate titles, we notice bias when describing the women who have recently come forward to discuss their sexual harassment. Fox News tends to use words like "allegations" and "investigations" whereas the San Francisco Gate highlights the words "women" and "harassment." Bias across the board is apparent when analyzing the words with the highest variance. We note the prominence of religious words and Christian subjects, clearly indicating the difference between sources that cover these topics and those that do not.

However, while these differences are apparent, the most valuable takeaway of this initial exploration was that the trained models achieved relatively low accuracy. If the goal is uniquely determining a news source given its contents, then 50% success cannot possibly be satisfactory. An important conclusion here is that, while these different sources evidently display dramatically different worldviews, word frequency analysis is not sufficient to capture the distinction. To achieve a successful discriminator, we need to employ more robust and complex linguistic analysis.

2.2 Manual Frame Matrix Construction

As described in the introduction, a more powerful way of approaching the subject of "bias" is to use Lakoff's notion of metaphorical framing. Specifically, Lakoff presents a framework in which we can assess various political statements, and lays the foundation to identify how some statements fall into the Strict Father worldview and others fall into the Nurturant Parent worldview. Discovering these different political frames in a particular corpus is a valuable indicator of the author's leaning and mindset.

Of course, because this technique is entirely novel, there does not yet exist a comprehensive example dataset of what frames for various issues may look like, not to mention the fact that frames may change tremendously over time and populations [22]. Therefore, a necessary first step in utilizing the concept of political framing is to fundamentally understand how they are constructed and manually curate a collection of Strict Father and Nurturant Parent frames. This is precisely what we do in this portion of the project.

2.2.1 Methods

For the task of understanding frames and creating an example of the relationships between Strict Father and Nurturant Parent frames, we break down the methods into two sections: identifying and categorizing frames, and constructing frame matrices.

Identifying and Categorizing Frames

Naturally, there are different frames for different topics — people speak of a "war on climate change" whereas they may gripe about "tax burden". These frames are indeed both very useful, but it is perhaps more helpful to compare frames within a single subject. As such, we specifically chose to analyze frames surrounding the issue of climate change, as there has been plenty of recent discussion on the issue and many different perspectives to address.

To find frames, we read any articles relating to climate change during Fall 2017 and Winter 2018, and took note of any interesting climate change frames. We then grouped these based on their subjects (e.g. the "war on ____" frame).

Constructing Frame Matrices

Using the list of grouped frames from the previous step, we then created frame matrices that break down how different frames evoke different feelings and represent different perspectives. For each group of frames, we can delineate between the frames that represent a Strict Father mindset and a

Nurturant Parent mindset. Additionally, we can divide the frames based on which side of an issue they represent — in the case of climate change, this means separating between those that accept climate change and those that deny it. This categorization allows us to visualize the relevant frames by subject for each worldview and opinion.

2.2.2 Results

The following tables show the constructed frame matrices for climate change. We identify a number of different subjects within climate change, and it is worth noting that not every subject necessarily has frames for all combinations of worldview and opinion.

2.2.3 Discussion

The frame matrices help to shed light on the how people from different worldviews discuss these issues differently. This is a fact that may be obvious — it would be unreasonable to expect people from completely different backgrounds to approach an issue with the same mindset.

A more interesting and valuable takeaway is that worldviews and opinions are not bound to each other; people of both the Nurturant Parent and Strict Father mindsets may or may not endorse a belief in climate change. This is contrary to the tendency to oversimplify the political spectrum and assume all people of each side share the same beliefs. Instead, we find that it is possible to connect with many people, regardless of their worldview, by talking about an issue in a way that resonates with them.

Frame	**Nurturant Parent**	**Strict Father**
nature	we are destroying nature	resources exist for humans
acceptors	–	"we must protect our god-given land" [23], "When I talk about Climate Change, I don't talk about science." [20]
deniers	–	"God gave us natural resources to use to make our lives better" [16]

Table 2.1: Manual Frame Matrix example

____ against climate change	war against; race against	war against
acceptors	"losing a war against climate change" [9], "And similar efforts to go after large industries for alleged damage to the public" [5]	–
deniers	–	"We've ended the war on beautiful, clean coal" [13]

clean ____	clean power plan; clean water rule; clean energy	clean coal
acceptors	"In the end, all of these methodological contortions are meant to obscure a very basic truth: that any 'savings' achieved by rescinding the Clean Power Plan will come at an incredibly high cost to public health and welfare. If the Trump administration is willing to make that trade, it should at least have the courage to admit it." [14], "They are closing because consumers are demanding energy from sources that don't poison their air and water, and because energy companies are providing cleaner and cheaper alternatives." [3]	"coal is ruining our land and giving the big guy all the benefits" [16]
deniers	–	"We've ended the war on beautiful, clean coal" [13]

response plan	mitigation	adaptation
acceptors	“if we unite together and vote for politicians that are committed to the mitigation of climate change and if we pressure legislators to adopt greener policies” [9]	"adapt by creating green technologies and casting out old ones" [20]
deniers	–	“If farmers and resort owners and mayors and naval planners all build with an eye toward how the future might change, then those changes as they arrive won’t be so harmful or expensive.” [4]

quality of life	all suffer from climate change	ruin living standards
acceptors	“We lose homes to climate change, but in much of the world families lose something far more precious: their babies. Climate change increases risks of war, instability, disease and hunger in vulnerable parts of the globe, and I was seared while reporting in Madagascar about children starving apparently as a consequence of climate change.” [10]	"climate effects will ruin property and land" [16]
deniers	–	“The extremists don’t care that ending our use of fossil fuels would dramatically reduce our living standards by making energy much more scarce and expensive, or that the poor would suffer far more than the wealthy.” [7]

responsibility	our job to protect	nature is in charge
acceptors	“Cities, too, are acting out of self-interest. By improving their air quality and becoming greener, cities turn into more attractive places to live and work. And where people want to live and work, businesses want to invest.” [3]	"we must defend our land from harm" [23]
deniers	–	“I’ve learned to be humble, respectful and vigilant in the face of nature’s power; to recognize that climate change can range from beneficial or benign to harmful or unbelievably destructive; and to understand that the sun and other powerful natural forces totally dwarf whatever meager powers humans might muster in attempting to control Earth’s climate and weather.” [7]

economic opportunity	creating green jobs	ruining old jobs (e.g. coal)
acceptors	–	–
deniers	“They simply think the road to salvation lies not through making do with less, but rather through innovation and the conditions in which innovation tends to flourish, greater affluence and individual freedom most of all.” [18]	–

is climate change bad?	threat to humanity	not a bad thing
acceptors	–	–
deniers	–	“Do we really know what the ideal surface temperature should be in the year 2100, in the year 2018?” [21], “once described carbon dioxide as a ‘harmless trace gas’ that was merely ‘plant food.’” [21]

3 Project Goals and Outline

3.1 Goals

Based on the preliminary exploration and research, the primary aim of this project is to develop a system that is able to automatically identify the Nurturant Parent and Strict Father frames for a given issue. We set these goals on the basis that: 1. We are able to reasonably separate news sources into Nurturant Parent and Strict Father buckets (as evidenced by the work and results of the Liberal-Conservative classification experiment) and 2. We are able to find frames for an issue and discern what side they fall on (as evidenced by the manual frame matrix construction).

The automatic discovery and identification of these frames is highly useful, especially in informing how we talk about issues to connect with a broader audience. In order to be compelling to multiple worldviews, we need to discuss topics in ways that resonate with each of them. As demonstrated, the alternative to such a tool is extremely arduous. It would involve manually reading through an immensely large corpus, taking note of potential frames and keeping track of their prevalence, and finally trying to group and categorize them based on arbitrary heuristics.

3.2 Outline

To achieve this goal, we break down the project into the following steps:

1. Data Collection
2. NP/SF Sorting
3. Collocation Extraction
4. Collocation Filtering
5. Key Sentence Collection
6. Frame Analysis

7. Output Synthesis

We elaborate on each of these in the following section.

4 Project Work

4.1 Data Collection

Collecting a sizable and satisfactory dataset for this project was immensely important. In order to determine frames across many perspectives and people, we needed to have a huge corpus that covered a variety of sources. To do this, we scraped news publications using the eventregistry.org API to find all recent articles pertaining to politics, and intentionally sourced from organizations with varied perceived biases. We ultimately compiled roughly 6000 articles, which totaled around 8 million words. Figure 4.1 shows the code used to collect the data.

```
import pandas as pd
import collections
from eventregistry import *

er = EventRegistry(apiKey = "d986a32c-55e1-4149-b841-1d1606fe2850")
conservative_sources = ["Breitbart",
                        "Fox News",
                        "National Review",
                        "The American Conservative",
                        "New York Post",
                        "Washington Times"]

liberal_sources = ["The New York Times",
                   "Washington Post",
                   "San Francisco Gate",
                   "HuffPost",
                   "The Economist"]

### SETTINGS ###

# 1. Set sources to either conservative_sources or liberal_sources
sources = liberal_sources

# 2. Set the output file name
output_file = 'NP_articles_2018-2019.csv'

q = QueryArticlesIter(conceptUri = er.getConceptUri("Politics"),
                      lang = "eng",
                      sourceUri = QueryItems.OR([er.getNewsSourceUri(source) for source in sources]),
                      dateStart = "2018-01-01",
                      dateEnd = "2019-04-01")

res = q.execQuery(er, sortBy = "date")

print("Number of articles:", q.count(er))
```

Figure 4.1: Data Collection Code

4.2 NP/SF Sorting

The next challenge was in dividing the data based on their worldview, specifically whether they fall into a Nurturant Parent or Strict Father mindset. We considered training a classifier to do this, the idea being that a model trained using words and semantics could distinguish between the two. However, we ruled out this approach due to the lackluster results of our initial Liberal-Conservative classification model. Furthermore, by using the generally perceived notions of how different organizations stand (e.g. Fox News is considered conservative while The Huffington Post is considered liberal), we concluded that we could achieve a reasonable and sufficient classification based on this simple heuristic. The following table shows all of the sources used and which worldview they were grouped in. Because we collected data with this division in mind, we had a roughly equal amount of articles and words in each of these two categories.

Nurturant Parent	Strict Father
The New York Times	Fox News
Washington Post	National Review
San Francisco Gate	The New York Post
Huffington Post	The Washington Times
The Economist	The American Conservative
–	Breitbart News

Table 4.1: Table of news sources

4.3 Collocation Extraction

4.3.1 Definition

A collocation is an arbitrary and recurrent word combination [2]. These words co-occur at a rate above the normal, and appear in common positions with respect to one another. We can use such occurrences of words as starting places for phrases that may contain valuable information, i.e. frames.

4.3.2 Algorithm

For the purpose of of extracting collocations from a corpus, we implement a modified version of the Xtract algorithm originally presented by Smadja in 1993 [15]. In particular, the original Xtract paper uses a part of speech tagger to increase accuracy and create phrasal, grammatical templates, whereas this implementation is self-contained and does not have these features. The method we implement is comprised of two major stages: 1) extracting significant bigrams and 2) combining multiple bigrams to form n-grams. The following sections outline the method in more detail.

Extracting Significant Bigrams

The goal of this step is to identify pairwise lexical relations and produce statistical information on pairs of words involved together in a corpus. There is strong evidence that most relations involving a word w can be retrieved by examining the words in the neighborhood of w within 5 words (without crossing sentence boundaries) [2]. Thus, we examine the text to discover word pairs that appear more often than expected by chance and appear in a relatively rigid way.

Step 1: Producing Concordances

In the first task, we take as input a given word w and output all of the sentences containing w. This is not as simple as it seems and is still an open problem. It is not fully correct to simply look for periods followed by whitespace as this fails in the case of titles (e.g. "Mr.") and acronyms (e.g. "N.B.A."). Our implementation could be improved by handling these cases more elegantly.

Step 2: Compile and Sort

In the second task, we take as input all of the sentences containing w and produce a list of words w_i with frequency information on how w and w_i co-occur. Specifically, we want to identify both the overall frequency of a word w_i as well as the relative frequency for each possible position of w_i (i.e. plus or minus 5 words from w).

Step 3: Analyze

In the third task, we take as input the list of words w_i with information on how they co-occur with w. We output significant word pairs along with some statistical information describing how strongly the words are connected.

First, we analyze the distribution of frequencies $freq_i$ for each of the collocates w_i and compute the average frequency $\overline{f}$ and the standard deviation σ. Then we replace $freq_i$ with a computation k_i, which is called the strength of a word pair.

$$k_i = \frac{freq_i - \overline{f}}{\sigma}$$

Second, we analyze the probabilities p_i^j of a word w_i occurring in each of the possible positions around w and produce the average $\overline{p_i}$ and variance U_i. We define this variance as the spread of a word pair. A word pair with a low spread means that the word w_i is relatively equally likely to occur in any position, whereas a pair with high spread means that the word w_i is much more likely to appear in some positions than others.

$$U_i = \frac{\sum_{j=1}^{10}(p_i^j - \overline{p_i})^2}{10}$$

Finally, we filter out word pairs to identify the most interesting word combinations. Specifically, we define thresholds k_0, k_1, and U_0, and filter out based on the following conditions:

$$C_1 : strength \geq k_0$$

$$C_2 : spread \geq U_0$$

$$C_3 : p_i^j \geq \overline{p_i} + (k_1 + \sqrt{U_i})$$

Condition one eliminates collocates that are not frequent enough. Condition two eliminates word pairs in which w_i may appear in any position relative to w as opposed to belonging to specific locations. Condition three pulls out the particular relative positions of the two words. As described by Smadja, the thresholds for each of these conditions must be experimentally determined, and we found that $(k_0, k_1, U_0) = (1, 1, 10)$ gave good results for our purposes.

From Bigrams to N-Grams

The goal of this stage is to combine the information from bigrams into larger collocations containing any number of words. These collocations more closely resemble real phrases and are therefore more informative.

Step 1: Producing Concordances

Identical to the first stage.

Step 2: Compile and Sort

Identical to the first stage.

Step 3: Analyze and Filter

The analyses in this step are much simpler than the first stage. We are only interested in percentage frequencies; for each of the possible distances from w, we analyze the distribution of words and only keep words if their probability is above a certain threshold T.

$$p(word[i] = w_o) > T$$

The output here gives us collocations that contain multiple words if they have a high probability of co-occurring.

4.3.3 Determining Interesting Words

The value of the collocation extraction is contingent on identifying good initial words. Depending on the input word, the results will be indicative of different topics, and possibly of specific frames. To explore collocations across a variety of topics, we chose to perform my experiments using the words “women”, “immigration”, and “climate”. The collocations we retrieve form a valuable baseline on which we can analyze and look for potential phrases and frames.

4.3.4 Collocation Results

The below output is an example of what the collocation extraction algorithm produces. For this example, we used the Nurturant Parent dataset and the input word "climate".

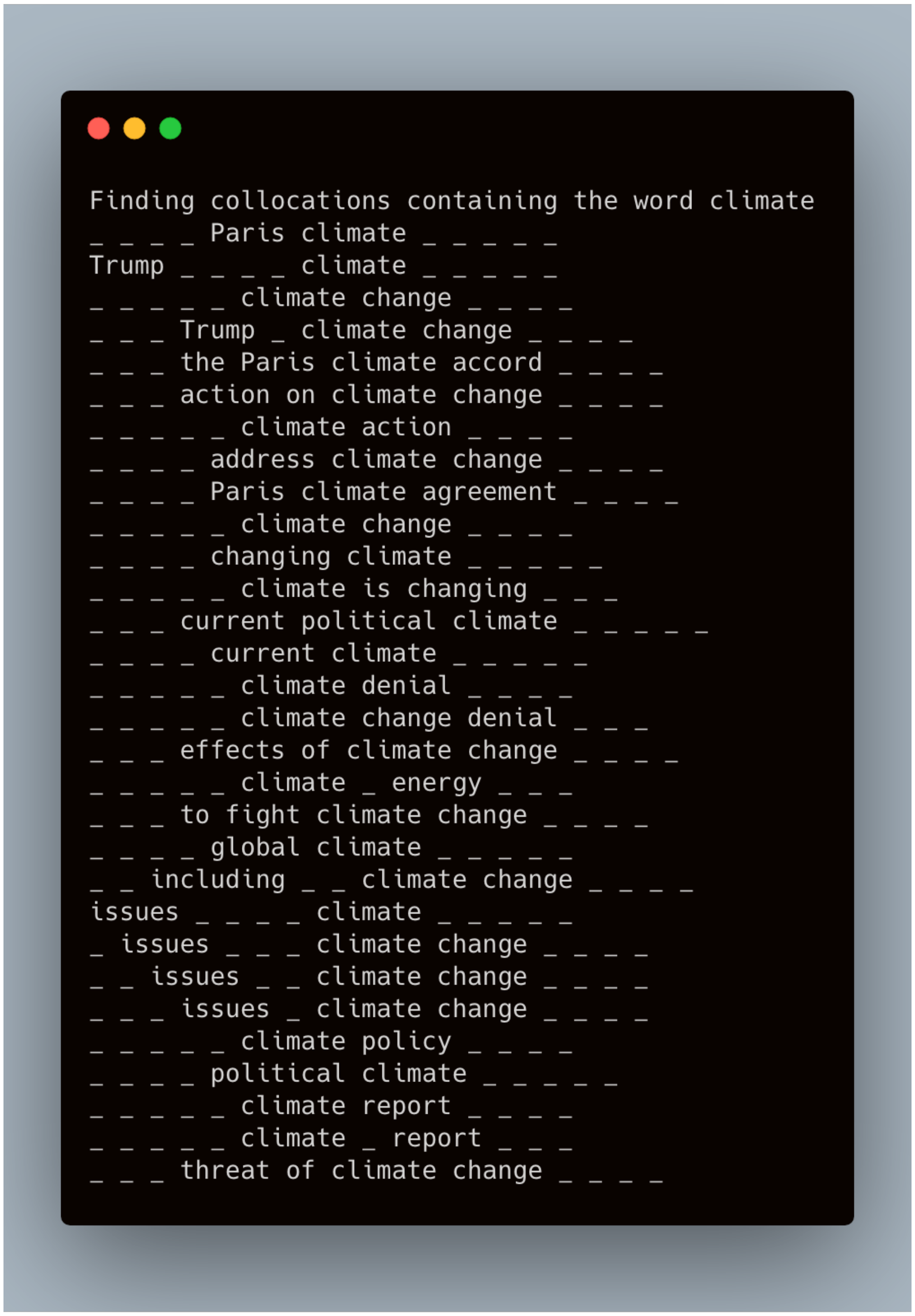
4.4 Collocation Filtering

The output of the collocation extraction is extremely useful, but without further discrimination we still have too large a dataset to use for frame identification. For each word, we identified 30-40 collocations for each of the Nurturant Parent and Strict Father worldviews. In order to distill this into a tractable number of frames, we need to cut the set down to a much smaller size.

This task boils down to the question of “How do you determine which collocations may represent metaphoric frames?” This remains an open question, and could spawn much more research on its own. For the purposes of this project, we simply needed a passable method as a first run. We considered using programmed heuristics to make the determination, for example, if collocations contain unique words relative to the others or if collocations contain loaded words. However, we decided that this method may not yield good results, and that given the number of collocations it was equally efficient to manually filter them.

Using a manual approach, we filtered each set of collocations down to roughly 20% of their original size.

Below are two example filtered collocation outputs for the input word "immigration". One output is for Nurturant Parent collocations; one output is for Strict Father collocations.



(a) NP "climate change" collocations

Figure 4.2: Collocation Extraction output

4.5 Key Sentence Collection

To identify frames, we need to analyze an entire sentence, as opposed to just the important words, as frames rely on the grammatical structure and semantic context. Thus, from collocations, we need to retrieve the sentences in which the interesting collocations are exactly present. Keep in mind that while the original bigrams needed very high frequency to exceed the imposed threshold, there are actually relatively few sentences that contain the exact n-grams since they were formed based on the likelihood of words from multiple bigrams existing together.

The below output is an example of the sentences collected from the Nurturant Parent dataset based on collocations containing the word "women".

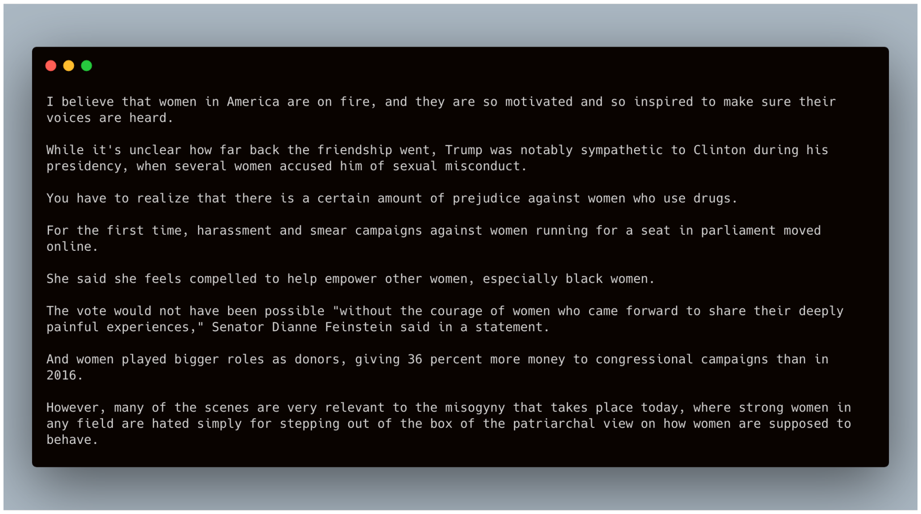
4.6 Frame Analysis

The next task is to pass these sentences into our frame analyzer in order to see the frame structure of each of the the sentences and identify metaphoric frames. To perform this analysis, we use the HLTC Automatic SRL C++ API [24]. We use specifically the sentences from our collocation

(a) NP "immigration" collocations

(b) SF "immigration" collocations

Figure 4.3: Filtered Collocation outputs



(a) NP "women" sentences

Figure 4.4: Excerpt of Key Sentence Collection output

extraction and sentence retrieval in order to limit the computation necessary, as well as filter the results to make identifying patterns much easier.

From the frame analyzer, we see grammatical and semantic relations emerge (e.g. “who did what to whom”) from which we can summarize the resulting frame structures. This allows us to see how frames manifest across each topic (recall that the experiments were performed with the input words “women”, “immigration”, and “climate”).

Formally, for a given collocation c, we retrieve the set of sentences S in which that collocation appears, and run the frame analyzer to collect the frame outputs F.

$$F = ASRL(s) \text{ for } s \in S \text{ where } contains(s, c) = True$$

Below are two example frame outputs for sentences containing the collocation "climate change". One is for a Nurturant Parent sentence: "The impact of climate change is increasingly visible: more violent storms, more wildfires and more severe droughts." The other is for a Strict Father sentence: "It’s like climate change, it can be whatever they want at any moment."

Through running the frame analyzer on many sentences, we collect frames that appear frequently throughout the corpus. For example, for the climate change issue, our preliminary tests revealed a common theme among the Nurturant Parent data of thinking about climate change as an aggressive agent, with phrases like "violent" and "hostile" appearing often. Meanwhile, in the Strict Father data, a theme of skepticism emerged, with phrases like "doubt" and "doesn’t stack up".


```xml
<?xml version="1.0" encoding="utf-8"?>
<value>
    <sent>
        <frame>
            <token>The</token>
            <pred type="TARGET">
                <token>impact</token>
            </pred>
            <arg type="A0">
                <token>of</token>
                <token>climate</token>
                <token>change</token>
                <token>is</token>
                <token>increasingly</token>
                <token>visible</token>
                <token>:</token>
                <token>more</token>
                <token>violent</token>
                <token>storms</token>
                <token>,</token>
                <token>more</token>
                <token>wildfires</token>
                <token>and</token>
                <token>more</token>
                <token>severe</token>
                <token>droughts</token>
            </arg>
        </frame>
        <frame>
            <token>The</token>
            <token>impact</token>
            <token>of</token>
            <pred type="TARGET">
                <token>climate</token>
            </pred>
            <token>change</token>
            <token>is</token>
            <token>increasingly</token>
            <token>visible</token>
            <token>:</token>
            <arg type="A1">
                <token>more</token>
                <token>violent</token>
                <token>storms</token>
                <token>,</token>
                <token>more</token>
                <token>wildfires</token>
                <token>and</token>
                <token>more</token>
                <token>severe</token>
                <token>droughts</token>
            </arg>
        </frame>
        <frame>
            <token>The</token>
            <token>impact</token>
            <token>of</token>
            <arg type="A1">
                <token>climate</token>
            </arg>
            <pred type="TARGET">
                <token>change</token>
            </pred>
            <token>is</token>
            <token>increasingly</token>
            <token>visible</token>
            <token>:</token>
            <token>more</token>
            <token>violent</token>
            <token>storms</token>
            <token>,</token>
            <token>more</token>
            <token>wildfires</token>
            <token>and</token>
            <token>more</token>
            <token>severe</token>
            <token>droughts</token>
        </frame>
        <frame>
            <token>The</token>
            <token>impact</token>
            <token>of</token>
            <token>climate</token>
            <token>change</token>
            <token>is</token>
            <token>increasingly</token>
            <token>visible</token>
            <token>:</token>
            <token>more</token>
            <token>violent</token>
            <token>storms</token>
            <token>,</token>
            <token>more</token>
            <token>wildfires</token>
            <token>and</token>
            <arg type="AM-MNR">
                <token>more</token>
                <token>severe</token>
            </arg>
            <pred type="TARGET">
                <token>droughts</token>
            </pred>
        </frame>
    </sent>
</value>
```


(a) NP "climate change" frames


```xml
<?xml version="1.0" encoding="utf-8"?>
<value>
    <sent>
        <frame>
            <arg type="A0">
                <token>It</token>
            </arg>
            <pred type="TARGET">
                <token>&apos;s</token>
            </pred>
            <arg type="A1">
                <token>like</token>
                <token>climate</token>
                <token>change</token>
            </arg>
            <token>,</token>
            <token>it</token>
            <token>can</token>
            <token>be</token>
            <token>whatever</token>
            <token>they</token>
            <token>want</token>
            <token>at</token>
            <token>any</token>
            <token>moment</token>
        </frame>
        <frame>
            <token>It</token>
            <token>&apos;s</token>
            <token>like</token>
            <arg type="A1">
                <token>climate</token>
            </arg>
            <pred type="TARGET">
                <token>change</token>
            </pred>
            <token>,</token>
            <token>it</token>
            <token>can</token>
            <token>be</token>
            <token>whatever</token>
            <token>they</token>
            <token>want</token>
            <token>at</token>
            <token>any</token>
            <token>moment</token>
        </frame>
        <frame>
            <token>It</token>
            <token>&apos;s</token>
            <token>like</token>
            <token>climate</token>
            <token>change</token>
            <token>,</token>
            <token>it</token>
            <token>can</token>
            <token>be</token>
            <arg type="R-A1">
                <token>whatever</token>
            </arg>
            <arg type="A0">
                <token>they</token>
            </arg>
            <pred type="TARGET">
                <token>want</token>
            </pred>
            <arg type="AM-TMP">
                <token>at</token>
                <token>any</token>
                <token>moment</token>
            </arg>
        </frame>
    </sent>
</value>
```


(b) SF "climate change" frames

Figure 4.5: Frame Analyzer outputs

4.7 Output Synthesis

The final task of the project is to combine the results from each of our analyses to determine frames across the Nurturant Parent and Strict Father worldviews as well as across the different perspectives on each political subject. The goal of this synthesis is to automatically create the frame matrices that we had previously constructed manually, as this structure allows for a clear viewing and understanding of the relevant frames.

For a given collocation, for both the Nurturant Parent and Strict Father sentences, we identify the semantic frames that follow along a common theme. These semantic frames groupings can be combined to highlight a particular frame for an issue.

Grouping frames along common themes is a challenge that can be solved in a number of ways. For this project, we identified possible themes and generated word sets that may represent them. For example, the previous section noted that Nurturant Parent frames revolved around an idea of climate change as an aggressive agent. So, we generate a set of candidate words W that possibly represent this theme, and find all the frames that contain one or more of these words.

Formally, we construct a group of frames G in which each frame contains at least one word w from the candidate set W.

$$G = \{(f \in F) \mid \exists w \in W \text{ such that } contains(f, w) = True\}$$

This leads us to output like the following.

Worldview	Strict Father
Frame	question validity of climate change
Ex. 1	"But [argument] simply doesn't stack up."
Ex. 2	"[Argument] has also expressed skepticism [argument]."
Ex. 3	"[Argument] did not recognize [argument]."
Ex. 4	"[argument] it can be whatever they want [argument]"

Table 4.2: SF climate output

Worldview	Nurturant Parent
Frame	climate change as a fight
Ex. 1	"[Argument] and attacks on scientists."
Ex. 2	"[Argument] is hostile to climate action"
Ex. 3	"[Argument] of aggressive climate action."
Ex. 4	"to reduce a clear and present danger, namely [argument]"

Table 4.3: NP climate output

Some frames manifest in both the Nurturant Parent and Strict Father datasets, which allows us to directly compare how similar themes get used in different ways. For example, both Nurturant Parent and Strict Father use framing around the idea of climate action being a monetary item, and use this to either justify or reject taking on the cost.

Combining the information from these mutual themes, we are able to recreate frame matrices like the ones developed manually.

Frame	**Nurturant Parent**	**Strict Father**
climate as a luxury good	can afford	too expensive
acceptors	**Ex. 1** "a country wealthy enough to [argument] but not wealthy enough to afford action on climate change", **Ex. 2** "[argument] affordable and clean energy"	–
deniers	–	**Ex. 1** "evidence that renewables do [argument] only much more expensively", **Ex. 2** "[argument] to pay the cost – which is the only sense that really matters"

Table 4.4: Synthesized Frame Matrix example

5 Analysis and Results

5.1 Algorithm

Summarizing the entire system, we can describe the algorithm as follows.

Algorithm 1 Frame Algorithm

```
procedure ExtractFrames
data collection:
    corpus ← getPoliticalArticles()
NP/SF sorting:
    corpus_NP, corpus_SF ← sort(corpus)
collection extraction:                          ▷ algorithm is described in section 4.3
    collocations_NP ← getCollocations(corpus_NP)
    collocations_SF ← getCollocations(corpus_SF)
collocation filtering:
    collocations_NP ← filter(collocations_NP)
    collocations_SF ← filter(collocations_SF)
key sentence collection:
    sentences_NP ← Dict < collocation : List() >
    sentences_SF ← Dict < collocation : List() >
    for c in collocations_NP do
        pattern ← createRegexPattern(c)
        for s in corpus_NP do
            if regex.match(s, pattern) then
                sentences_NP[c].add(s)
    for c in collocations_SF do
        pattern ← createRegexPattern(c)
        for s in corpus_SF do
            if regex.match(s, pattern) then
                sentences_SF[c].add(s)
```

```
frame analysis:
    asrl_NP ← Dict < collocation : Counter() >
    asrl_SF ← Dict < collocation : Counter() >
    for c in sentences_NP.keys() do
        for s in sentences_NP[c] do
            frames ← ASRL(s)
            asrl_NP[c].add(frames)
    for c in sentences_SF.keys() do
        for s in sentences_SF[c] do
            frames ← ASRL(s)
            asrl_SF[c].add(frames)
output synthesis:
    frames_NP ← Dict < theme : List() >
    frames_SF ← Dict < theme : List() >
    candidateThemes ← identifyCandidateThemes()
    for t in candidateThemes do
        candidateWords ← identifyCandidateWords()
        for f in asrl_NP do
            for w in candidateWords do
                if f.contains(w) then
                    frames_NP[t].add(f)
        for f in asrl_SF do
            for w in candidateWords do
                if f.contains(w) then
                    frames_SF[t].add(f)
finish:
    return frames_NP, frames_SF
```

5.2 Qualitative Analysis

In this section, we analyze our method and compare it to possible alternatives. The algorithm implemented in this project is a first-pass functional implementation of a automatic frame identification. We set the goal of recreating the frame matrices that we initially created manually, and committed our focus to successfully retrieving political frames. However, with this target in mind, we did not focus at all on efficiency, and some parts of the algorithm require manual input.

Specifically, we require manual work for identifying possible frame themes and generating a list of candidate words that represent that theme. In addition to increasing the total time necessary for the algorithm, this method has the negative effect of introducing a subjective component. We rely on a judicious and nonpartisan creation of themes and candidate words, but it would be naive to assume that such a strategy is perfect. One alternative to this approach is to use sentiment analysis to identify the strongest and most powerful words given a set of sentences. For example, after generating the set of sentences containing the collocation "climate change", we could run sentiment analysis on the data and find the frequent and powerful words. In our project, we decided against this approach as we were not sure if sentiment analysis would yield informative and interesting words, or simply yield words that reflect the overall affect of a sentence (i.e. "happy" words from "happy" sentences).

Regarding efficiency, our algorithm loops over the data many times to filter it down. First, the entire corpus is looped over in order to extract collocations. Then the collocations are looped over for filtering. Then we loop over both the corpus and collocations to collect key sentences. Then we loop over the sentences to identify semantic frames. And finally we loop over both these sentences and the candidate words to gather our final frames. Ideally, we would be able to merge some of these loops into a single pass in order to improve the computational efficiency of the algorithm. If not that, we would at least introduce parallelization in order to speed up each of the steps.

5.3 Quantitative Analysis

In this section, we aim to analyze our method quantitatively. A significant challenge in doing this is that there is no existing test set available — because we are developing an algorithm to produce an entirely novel output, we have nothing to judge against. In order to create a test set, we must randomly sample N sentences from our corpus, and manually identify the frames as done previously. However, this presents a complication: because our algorithm relies in part on manual input, our perception will be skewed based on the data in the test set. Therefore, our approach for quantitative analysis is as follows.

5.3.1 Experimental Setup

First, we choose a subject to test on, and then identify possible themes for frames in that subject. For each of these themes, we create a set of candidate words that may represent those themes. Next, we randomly sample sentences from the sentences dataset for that subject, and manually identify frames. Finally, we run our algorithm on the randomly sampled dataset to see the frame outputs, and compare against our manual results.

Below we show the candidate word sets created for our experiments.

Climate Change	
Theme	Candidate Words
War	war, against, fight, battle, resist, stand, lose
Money	cost, pay, afford, money, expensive, invest, price, economic, fund
Clean	renewable, energy, green, coal, power, environment
Responsibility	protect, defend, care, keep, choice, vulnerable
Skepticism	doubt, skeptical, real, certain, deny, science
Higher Power	nature, god, force, natural

Table 5.1: Candidate words for test

5.3.2 Experiment Results

For the following test we use the subject "climate change" and $N = 50$.

Theme	# True Positives	# False Positives	# Actual Positives
War	3	3	5
Money	4	2	7
Clean	3	2	2
Responsibility	3	0	9
Skepticism	0	0	2
Higher Power	1	0	1

Table 5.2: NP Test results

$$Precision_{NP} = \frac{14}{21} \approx 0.667$$

$$Recall_{NP} = \frac{14}{26} \approx 0.538$$

$$f\text{-}score_{NP} \approx 0.596$$

Theme	# True Positives	# False Positives	# Actual Positives
War	3	5	6
Money	1	2	4
Clean	0	1	1
Responsibility	2	3	5
Skepticism	5	3	9
Higher Power	0	1	0

Table 5.3: SF Test results

$$Precision_{SF} = \frac{11}{26} \approx 0.423$$

$$Recall_{SF} = \frac{11}{25} \approx 0.440$$

$$f\text{-}score_{SF} \approx 0.431$$

We note that the method performed much better on the NP dataset than the SF dataset. This is most likely due to the candidate word sets created. As noted in the qualitative analysis section, the generation of the candidate word sets introduces a level of subjectivity. These experiments were run entirely by individuals of the NP mindset, and therefore the candidate words a likely a reflection of that.

Additionally, we see contrasting levels of precision and recall. This too can be attributed to the candidate word sets we created. With thorough candidate word sets, the system is able to pick up on all of the possible frames, missing very few because the candidate words are able to cover everything. In other words, a thorough candidate word set results in high recall. However, with a larger candidate word set, the system will also have a lower precision, as there are more potential cases for the system to include a frame when it should not. Below we describe some example cases.

A simple true positive is from the NP sentence "We have no choice but to address climate change, or it will address us." The framing around the word "choice" is clear here, and the system appropriately captures this instance.

A missed element is apparent in the sentence "Australia's crisis shows that, paradoxically, demand for action on climate change is likely to grow, both to address pollution, the underlying cause of climate change, but also the consequences of inaction." The frame "consequences of inaction" is powerful and noteworthy, but the system does not pick up on it none of the candidate words are present.

5.3.3 Resulting Matrix

From the results of our experiments, we can also create a frame matrix. To do this, we simply gather the true positives for a given theme for both Nurturant Parent and Strict Father, and then organize them appropriately.

Frame	**Nurturant Parent**	**Strict Father**
Money	–	–
acceptors	**Ex. 1** "After months of negotiations, a rift within the party escalated last weekend over an energy proposal, which was meant to reduce electricity prices and address climate change by cutting emissions." **Ex. 2** "She was, instead, talking about the economic impact of policies to fight climate change, which she conceded would adversely affect some industries even as it helped others." **Ex. 3** "A country wealthy enough to offer corporate tax cuts and subsidies to the coal industry, but not wealthy enough to afford action on climate change or high-quality care for the elderly." **Ex. 4** "* News analysis: As the Trump administration continues to cut environmental regulations, it is widely expected to discount or ignore a major government report issued on Friday that details the economic effects of climate change."	–
deniers	–	**Ex. 1** "But the truth is that 'climate change' - at least as perceived by the IPCC - is bunk and all that expenditure (which, added up, amounts to a sum greater than the entirety of global GDP) would be a complete waste of money"

Table 5.4: Experiment Frame Matrix

6 Discussion

The project on the whole reveals extremely interesting results regarding how political issues are discussed across different perspectives and worldviews. Just from the manual frame matrix construction, it is evident that different people speak of different subjects differently. More importantly, this demonstrates that different people feel passionately about issues only when they are presented in a way that speaks to them. This relates back to Lakoff's fundamental theories that people use language that carries important weight for them specifically, but that exactly what language has this effect varies broadly from Nurturant Parent to Strict Father mindsets, and additionally varies from one person to the next.

This finding has implications for how different subjects should be discussed to captivate and inspire the largest audience possible. With climate change, for example, Nurturant Parent frames revolve around the notion of preventing harm from coming to the environment, which is portrayed as helpless against human forces. In contrast, Strict Father frames that push for climate action focus more on the notion of protecting our land and punishing those who harm or disrespect nature. This difference is reminiscent of some of the Nurturant Parent vs Strict Father viewpoints that Lakoff discusses, including the tendency to offer aid and the desire to adhere to tradition and values.

Even before the final frame outputs, the intermediate results are highly informative on their own. One interesting area is in comparing how much discussion a topic receives from Nurturant Parent versus Strict Father sources. For collocations counting the word "climate", we extracted three times as many from the Nurturant Parent corpus than from the Strict Father corpus (30 collocations vs 10 collocations). This reveals how much attention the different sources give to this issue, and is therefore a reflection of the priorities and bias of some sources over the others.

Additionally, we notice that identifying leaning through collocations is much more revealing and interesting than using word analysis alone. For example, in the collocations for the word "women", while there are many collocations that overlap or are similar, we notice a stark contrast in the content of some collocations between Nurturant Parent and Strict Father. For Nurturant Parent, we see frames like "strong women", "women _ encouraged", and "opportunities _ women". For Strict Father, we see frames like "allegations _ _ _ _ women", "protect women", and "vulnerable women". Again, these differences highlight some of the different viewpoints discussed by Lakoff, specifically a sense of equal opportunity and a responsibility for defending one's family.

In the final frame outputs, we see both positives and negatives. On one hand, the system does a reasonable job at capturing many of the frames in the corpus, and helps guide us to a number of interesting themes and language uses throughout the data. On the other hand, we see that a critical point of the system is in the strong identification of themes and candidate words, and this creates a bottleneck that is very likely imperfect. To combat this issue, we must determine a means to automatically generate sets of candidate words, so that there is no human error involved.

7 Further Work

This project is highly exploratory; nothing similar to this has been done before and much of the implementation and design was created as a functional first pass. There are a number of areas for improvements in efficiency and optimality. The data collection methodology, for example, can be expanded to include more text than simply news articles. Other interesting sources could include interview transcripts, forum posts, or short essays. The classification of Nurturant Parent and Strict Father data could also be improved upon. For this project, we simply applied a heuristic based on the source's colloquial leaning, but a more robust and accurate method would likely include training a model. One might also improve collocation extraction by adding a part of speech tagger, and improve collocation filtering by using a model or programmed heuristics as opposed to a manual review. All of these changes would work to enhance the success of the system, although the fundamental capabilities would remain the same.

An area of work not scoped in this project would be to develop a user-in-the-loop frame matrix construction tool, in which the outputs of this project are combined with user input to create highly accurate frame matrices. While the frames and collocations that this system identifies may be a satisfactory baseline, a truly successful system would require some human oversight to ensure the frames are properly categorized and filtered. Ideally this would be a final interface in which the frame matrices are displayed and made mutable.

Ultimately, we hope that the discoveries and work from this project inform future work in the political language space. As suggested in the discussion, knowing the frames that people use to discuss particular issues allows one to more effectively communicate to different groups of people. This would prove massively beneficial for the purposes of galvanizing an audience or voting group. People may develop different worldviews, but that does not necessarily mean they disagree on every issue.

Bibliography

1. D. M. Amodio. *Neurocognitive Correlates of Liberalism and Conservatism.* URL: Nature%20Neuroscience, %20vol.%2010,%20no.%2010,%20Sept.%202007,%20pp.%201246%E2%80%931247.,%20doi: 10.1038/nn1979.
2. M. Benson. *Collocations and general-purpose dictionaries.* URL: International%20Journal%20of%20Lexicography,%203(1),%2023-35. (visited on 1990).
3. M. R. Bloomberg. *Climate Progress, With or Without Trump.* URL: www.nytimes.com/2017/03/31/opinion/climate-progress-with-or-without-trump.html (visited on 03/31/2017).
4. O. Cass. *Trump Is Wrong on Climate Change.* URL: www.foxnews.com/opinion/2017/03/29/trump-is-wrong-on-climate-change.html (visited on 03/29/2017).
5. N. C. Cepeda. *Focus: Novel Legal Strategy Underpins Sea Level Rise Lawsuits against Oil and Coal Companies.* URL: www.sandiegouniontribune.com/news/environment/sd-me-climate-change-legal-20170723-story.html (visited on 07/23/2017).
6. D. Domke. *News Media, Racial Perceptions, and Political Cognition.* URL: Communication%20Research, %20vol.%2026,%20no.%205,%201%20Oct.%201999,%20pp.%20570%E2%80%93607.,%20doi: https://doi.org/10.1177/009365099026005003.
7. P. Driessen. *Don't Blame Climate Change: Hurricanes Have Always Been and Will Always Be with Us.* URL: www.foxnews.com/opinion/2017/09/16/dont-blame-climate-change-hurricanes-have-always-been-and-will-always-be-with-us.html (visited on 09/16/2017).
8. T. Groeling. *Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News.* URL: Annual%20Review%20of%20Political%20Science,%2028%20Feb.%202013,%20doi: 10.1146/annurev-polisci-040811-115123..
9. S. Huang. *Losing a War Against Climate Change.* URL: www.huffingtonpost.com/entry/losing-a-war-against-climate-change_us_59928db9e4b0caa1687a6349 (visited on 08/21/2017).
10. N. Kristof. *We Don't Deny Harvey, So Why Deny Climate Change?* URL: www.nytimes.com/2017/09/02/opinion/sunday/hurricane-harvey-climate-change.html (visited on 09/02/2017).
11. G. Lakoff. *The All-New Don't Think of an Elephant!: Know Your Values and Frame the Debate.* URL: Chelsea%20Green%20Publishing,%202014..
12. D. Mladenic. *Learning How To Detect News Bias.* 2015.

13. B. Plumer. *What 'Clean Coal' Is - and Isn't.* URL: www.nytimes.com/2017/08/23/climate/what-clean-coal-is-and-isnt.html?rref=collection (visited on 08/23/2017).

14. R. L. Revesz. *The E.P.A.'s Smoke and Mirrors on Climate.* URL: www.nytimes.com/2017/10/09/opinion/environmental-protection-obama-pruitt.html (visited on 10/10/2017).

15. F. Smadja. *Retrieving collocations from text: Xtract.* URL: Comput.%20Linguist.%2019,%201%20(Mar.%201993),%20143-177. (visited on 1993).

16. B. Staff. *Does God Really Care How We Use Our Natural Resources?* URL: billygraham.org/answer/does-god-really-care-how-we-use-our-natural-resources/ (visited on 06/01/2004).

17. K. E. Stanovich. *Myside Bias, Rational Thinking, and Intelligence.* URL: Current%20Directions%20in%20Psychological%20Science,%20vol.%2022,%20no.%204,%202013,%20pp.%20259%E2%80%93264.,%20doi:10.1177/0963721413480174..

18. B. Stephens. *Apocalypse Not.* URL: www.nytimes.com/2018/02/08/opinion/environment-oil-scare.html (visited on 02/08/2018).

19. A. A. Strickland. *Motivated Reasoning and Public Opinion.* URL: Journal%20of%20Health%20Politics,%20Policy%20and%20Law,%20vol.%2036,%20no.%206,%202011,%20pp.%20935%E2%80%93944.,%20doi:10.1215/03616878-1460524..

20. A. D. Thaler. *When I Talk about Climate Change, I Don't Talk about Science.* URL: www.southernfriedscience.com/when-i-talk-about-climate-change-i-dont-talk-about-science/ (visited on 01/06/2017).

21. N. Visser. *EPA's Pruitt Suggests Global Warming May Not Be A 'Bad Thing' For Humans.* URL: www.huffingtonpost.com/entry/scott-pruitt-global-warming-good_us_5a7ba9bce4b044b3821922dc (visited on 02/08/2018).

22. M. Wolfgang. *A stochastic case frame approach for natural language understanding.* URL: Proceeding%20of%20Fourth%20International%20Conference%20on%20Spoken%20Language%20Processing.%20ICSLP'96.%20Vol.%202. (visited on 01/01/1996).

23. Wolsko. *Framing Discourse around Conservative Values Shifts Climate Change Attitudes.* URL: oregonstate.edu/ua/ncs/archives/2016/apr/framing-discourse-around-conservative-values-shifts-climate-change-attitudes (visited on 10/05/2017).

24. D. Wu. *HLTC Automatic SRL C++ API.* URL: http://hltc.cs.ust.hk/asrl_api_documentation/index.xhtml (visited on 05/10/2019).

25. S. K. Yeo. *News Selectivity and Beyond: Motivated Reasoning in a Changing Media Environment.* URL: Publizistik%20Und%20Gesellschaftliche%20Verantwortung,%202015,%20pp.%2083%E2%80%93104.,%20doi:10.1007/978-3-658-04704-7_7..