

MACROANALYSING OF UNITED STATES PRESIDENTIAL SPEECHES

Digital Humanities and Information Technologies Final year report

Harry O'Callaghan – 120369996

Contents

Abstract.....	4
Introduction	5
Chapter 1	7
Literature Review	7
1.1 Political speeches and Natural Language Processing	7
1.2 Collocation Analysis.....	8
1.3 Keyword Extraction	10
1.4 Sentiment Analysis	11
1.5 Craig Zeta Analysis.....	13
Chapter 2	13
Tools and Methods.....	13
2.1 VScode and natural Language Toolkit (NLTK).....	13
2.2 Collocation Analysis.....	14
2.3 Keyword Extraction	16
2.4 Sentiment Analysis	17
2.5 Craig Zeta Analysis.....	20
Chapter 3	21
3.1 Analysis of the first research question.....	21
3.1.1 Collocation Analysis	22
3.1.2 Keyword Extraction	26
3.1.3 Sentiment Analysis	30
3.2 Conclusion.....	33
Chapter 4	34
4.1 Analysis of the second research question.....	34
4.1.1 Collocation Analysis	34
4.1.2 Keyword Extraction	39
4.1.3 Sentiment Analysis	44
4.2 Conclusion.....	47
Chapter 5	48
5.1 Analysis of final research question	48
5.1.1 Words Preferred (Democratic)	48
5.1.2 Words Preferred (Republican).....	51
5.1.3 Comparison and contrasting.....	53
5.2 Conclusion.....	55

Chapter 6	56
Overall Conclusion	56
Bibliography.....	58

Abstract

Macroanalysis of these presidential speeches investigates the change in key points over time between cold war presidents and modern-day presidents and the difference between both political parties' key points. The report aims to understand the chosen political speeches using Natural Language Processes and Craig Zeta analysis to allow for further research using these methods in the analysis of political speeches. The methods that are used throughout the project are collocation analysis, keyword extraction, sentiment analysis and Craig Zeta analysis, since they are the best fit for this project. The results indicated a steady change in key points from cold war presidents and modern-day presidents in their inaugural speeches and certain speeches around shared global and internal issues. The craig zeta analysis also showed how both parties present themselves to the American people. These tools and analysis techniques shows how powerful they are when analysing political speeches and allow for a great understanding of the speeches. To further understand how the analysis of the presidential speeches was done, you can find my code in my digital artifact here at this link <https://github.com/HarryOC/Harry-O-Callaghans-Final-Year-Project-Code>.

Introduction

In this report I aim to answer three questions, the first being “Using Natural Language Processes (NLP), how have United States Inaugural speeches key points changed from cold war presidents to modern day presidents” with the second question being “How do Cold War presidents and modern-day presidents tackle shared global issues or internal issues in their speeches?” and finally “What key phrases emerge from Democrat and Republican presidential speeches when analysed using Craig Zeta analysis”. I have chosen this topic as I have always had a great interest in United States politics, which was amplified by taking a minor subject in politics over the course of my years in college paired with my interest in textual analysis. I believed it would be a great place to focus my final year project. With this both in mind after spending some time reading and listening to old presidential speeches, it became clear that this is something I could research in my final year project and truly understand how key points in US presidential speeches have changed and or stayed the same from the period of 1946 to 2018. I hope to contribute to political science using digital humanities and show how digital humanities in the form of text analysis tools can be used to get insights into political speeches and show how both fields can work in tandem to generate new ideas and new research questions to advance both areas.

For the scope of this essay, I have chosen to narrow down the presidential speeches to six United States presidents as I could not analyse all US presidential speeches in the timeline of this project. which are, from the democratic party Harry S. Truman, John F. Kennedy, and Barack Obama. From the republican party I have chosen Ronald Reagan, Gorge W. Bush, and Donald Trump. I have gathered these speeches from Kaggle, where I found a dataset made by Kristof Boghe (Boghe, 2020) which collected presidential speeches in the United States from 30/12/1798 to 13/03/2020 which encompasses all the presidential speeches I need for this project. After analysing the excel sheet, I picked out the speeches that I could analyse using NLP then put them into a new excel file after which I then put each of the speeches into their own .txt file which allowed me to clean the data and remove all unwanted written terms such as “Applause” which appears throughout each speeches text.

For the first question, this report will look deep into all their inaugural speeches using NLP to identify the key points from their speeches. In the second question of this research project, this report will compare speeches from different presidents on similar global issues or internal issues. The speeches I will compare will be from important moments in US history and will be matched with speeches from today's presidents that focus on similar issues. This

way, we can see how current presidents handle past key issues that echo into the future. The speeches I have chosen to compare are, Trumans “Truman Doctrine” speech (March 12, 1947) to Goerge Bush’s remarks on the Freedom in Iraq and the Middle east (November 6, 2003), along with comparing John F. Kennedy Speech on the build-up of arms in Cuba (October 22, 2962) to Barack Obama’s remarks on the people of Cuba (March 22, 2016) and finally comparing Ronald Reagan speech to the nation on the campaign against drug abuse (September 14, 1986) to Donald Trump’s Remarks on combatting the Opioid Crisis (March 19. 2018). In this comparison, party lines will not matter as it will be looking at cold war presidents against modern day presidents. For the final question all the Inaugural speeches and all speeches declared above, will be put into two categories which are the democratic party presidents and the republican party presidents and compared against each other using the Craig Zeta analysis technique to identify party presidents key points through cold war presidents and modern-day presidents.

Chapter 1

Literature Review

1.1 Political speeches and Natural Language Processing

To understand the aims of this paper, one must first understand the power of speeches.

Speeches are one of the most powerful vocal tools we have. They have been a practice for many decades and are now a staple of politics. One such political system that uses political speeches as a main pillar are the United States of America. For a United States president, their time leading up to their election, is controlled by speeches including debates against candidates and speeches in specific states important for an election. Once a president arrives at office, there are a few important speeches of their term. The inaugural address at the start of their term and the yearly state of the union address alongside the many speeches given on a range of topics important for the United States. As stated by Matthew Eshbaugh-Soha (Eshbaugh-Soha, 2008) presidential speeches are to signal many preferences by the bureaucracy the presidents use their speeches to affect and outline policy. From looking at this, we can clearly get a sense of how important this is for the bureaucratic process in the United States and how the presidential word, is in many cases the word of policy in the country. This is backed up by Jessica L. Hughes (Hughes, 2009) that recent research concludes that presidents can set policy agenda in line with what they believe important.

For much of the research carried out on the political speeches, many of it is done using Natural Language processing (NLP) which as described by Dorothy Young Brockopp (Brockopp, 1983) is a Nueuro-Lingustic programming which is a new approach to understanding the process of human communication. This article is by far outdated but it shows that this is a practice that by now is a few decades old. It has advanced much since then. Technologies based on NLP has become more widespread, with examples such as phones and computers that support predictive text and handwriting recognition (Bird, Klien, & Loper, 2009) this shows that the growing digital word, is advancing our technologies in how we understand NLP. This leads to the question of whether, this can be used in political speech analysis by using one of the technologies that have been built around NLP which is Natural Language Toolkit (NLTK). As described by the NLTK website (NLTK, 2023), it is a platform for building Python programs to work with human language data. Its very design is for textual analysis. Its purpose is for all text, so does this include political speeches, as stated here by Katre (Katre, 2019), which from his discussion he concludes that Natural Language

Processing techniques are extremely helpful when understanding and interpreting the transcript of political speeches. With NLP being a very well research area with numerous works on all kind of text and many focusing on political speeches using NLTK and other computational methods, that there are very few gaps in this area of research and with the understanding that it is a sound way of carrying out research on Political speeches.

NLP is a widefield, with many ways to analyse text. My decision to use Collocation Analysis, Keyword Extraction, Sentiment analysis and Craig Zeta analysis was made as I believed they were the best for my research question but as it is a widefield there were many processes I could have used. In the NLP community, there are many questions about what is considered as NLP, for example this work, concludes that the aim of their poll was to establish what experts from within the NLP community consider to be NLP. (Rijk, Derks, & Grimley , 2019), clearly from this, it shows that it's such a wide community that so many different opinions would be centred around what belongs to NLP, clearly showing that it is a wide field for research. This is backed up by work done by Davide Falessi, Giovanni Cantone, and Gerardo Canfora with their title "A comprehensive Characterization of NLP Techniques Identifying Equivalent Requirements" which they use 242 NLP techniques to identify equivalent requirements in the context of their research. (Falessi , Cantone, & Canfora, 2010). With so many techniques, it shows how broad it is for research and gives so many approaches to answering the questions researchers would pose.

1.2 Collocation Analysis

Collocations will be a major part of my analysis of my chosen speeches. As described by McKeown and Radev (McKeown & Radev, 1997), that a collocation, is a lexical phenomenon that has linguistic and lexicographic as well as utility for statistical natural language paradigms. They continue by stating that this is exactly why it is featured in many statistical approaches to natural language processing, with an example, bellow with Fig 1. As explained by Hamad (Hamed, 2021) collocation is the frequent co-occurrence of words within a certain distance recognised either side of the specified focal word or node. Some of collocations that are a non-fixed combination of words, base and a collocutor are if Fig 2 below. This makes it clear as to why this approach would be taken in NLP research questions and can be seen placed in works by many researching political speeches. This work by Dalia Hamed (Hamed, 2021) aims to use a corpus-assisted study to investigate keywords and their collocations in the USA presidential discourse from presidents starting with Bill Clinton all the way to Donald Trump. Hamed would use collocates and keywords to reach this aim.

Hamed describes collocations as the frequent co-occurrence of words within a certain distance. The results from Hameds work clearly shows the usefulness in terms of corpus research on speeches. For example, “The adjective “great” collocates with “people, American, job, doing, country”, which do not specify a goal set to be fulfilled.” (Hamed, 2021). This shows that for this research, they decided to collocate words that are in proximity, and ignoring the rest of the sentence that the word would be collocated to. Which shows the direction this research decided to take. This leads to a smaller dataset of words around a collocate, ensuring that you are looking at it through a lens almost and you are missing the stuff around the lens.

For collocations, especially in speeches, it should be a larger gram used to be able to understand the content of the speech in greater detail, so that words that might be connected in a bigram might be connected in a larger gram that would give more detail on the content of the speech. Dalia (Hamed, 2021) still concludes that the application of keywords and the collocations are believed to be tools, that give credible analysis and describe the corpora in a reliable manner. We see that Dalias work, gives credible research on the topic, and the chosen research topics. Collocations are a well-known and documented part of NLP, it requires very little research on the usefulness of this an NLP process and as it is a pillar of NLP, it is highly researched which leaves very few gaps in the knowledge of collocations. The only thing that when working with collocations, is the thought of how many N-Grams one should use when collocating speeches or any text. An n-gram is “N-Grams or Collocations are nothing but two or more words that tend to appear frequently together” (Yadav, 2021). This allows us to understand, it’s just the way that we decide how many words we wish to connect, with a bigram being two and a trigram being 3 and so on. The higher the gram the better the insight into the text. This idea that collocations are useful is backed by Seretan, who says “We believe that many language processing tasks may considerably benefit from the approach of multi-word collocation extraction using linguistic constraints.” (Seretan, Nerima, & Wehrli, n.d.) Overall, when it comes to collocations, its strong way to analyse text and comes down to the indivual when deciding how many N-Grams one wishes to use. In much simpler terms, it gathers the words with proximity that can then give an idea of the words used and key sentences in the gathered text.

Idioms	Collocations	Free Word Combinations
to kick the bucket	to trade actively	to take the bus
dead end	table of contents	the end of the road
to catch up	orthogonal projection	to buy a house

Example of a collocation (FIG 1) (McKeown & Radev, 1997)

- verb + noun as in: express admiration;
- adjective + noun, as in: serious consequences;
- noun + verb, as in: a problem persists;
- noun + noun, as in: job market;
- adverb + adjective, as in: deadly serious; and
- verb + adverb, as in: (to) sleep soundly.

Non-Fixed combination of words, a base and a collocator (Fig 2) (Hamed, 2021)

1.3 Keyword Extraction

Keyword extraction was mentioned in the above paragraph, and much like collocation it will have an important role to play in this report. Slobodan Beliga best describes Keyword Extraction by saying That Keyword Extraction is a task that automatically identifies a set of the terms that best describe the subject in the document (Beliga, n.d). By the description of the Cambridge dictionary a key word is described as “a word that you type into a computer so that the computer will find information that contains that word” (Keyword , n.d.), this gives the impression, that a keyword is a way of fishing information by using a specific process. Which is what is happening, in keyword extraction, by singling out keywords, you then can find the subject of the document by analysing the keywords in the text piece.

Keyword Extraction is a Natural Language Process used by many when researching text pieces, for example a study done on Hillary Clinton and Donald Trump’s Linguistic styles done by Xueliang Chen. Yuanle Yan and Jie (Chen, Yan, & Hu, 2019) they us keyword extraction as a main pillar of their research. They get thirty of the top words used by Clinton and 30 top words used by Trump. They create a graph to give the number ranking from 1-30, the keyword itself, and the frequency of the word, which can be seen bellow in Fig 3. From a quick glance of their work, it shows the rhetoric and the key points of both politicians as it is very clear, what they discuss in most of the speeches. They discuss their results and conclude that Hillary Clinton was more inclined towards rational discussions of public policy while Trump was adept to appealing to voters’ emotions (Chen, Yan, & Hu, 2019). This research is a great indication of the usefulness of keyword extraction when it comes to analysing text.

One thing, that I would argue, is that they seem to use the raw frequency of the words and not the relative frequency, they do not indicate the size of the corpus and the lengths of the texts they use for their analysis so if a certain word appears more in a Clinton speech, but it is a much longer speech than Trumps, but Trump might have a higher percentage use of the word, then it is clear that it is just that Hillary's speech was much longer and statically Trump would have used the word more frequently. By using the relative frequency, it would give a greater insight into the keywords as the percentage of how much of it appears to be analysed, rather than the number of times it is said in a speech. In the end the choice between relative and raw frequency is up to the researcher and how they wish to work and present their data. Nonetheless, keyword extraction is perfect way to understand the subject matter of speeches and of any textual document. It is a highly researched area and greatly defined and as concluded by Beliga, keywords provide a compact representation of a documents content (Beliga, n.d) The hope of this research is to build on this work done by many researchers and come to understand myself the usefulness in keyword extraction to get the subject matter from the speeches.

Number	Keyword	Frequency	Keyness	Number	Keyword	Frequency	Keyness
1	Clinton	286	831.004	16	rights	89	159.791
2	he	938	391.441	17	kids	110	153.023
3	his	354	356.325	18	Trump	446	151.121
4	Donald	308	320.411	19	sure	149	144.434
5	together	257	260.637	20	America	450	132.135
6	my	520	255.01	21	can	696	126.088
7	work	376	239.52	22	someone	60	122.249
8	college	122	236.789	23	women	144	121.735
9	economy	152	209.479	24	com	45	121.65
10	families	154	206.521	25	hard	174	118.46
11	young	137	187.857	26	dad	48	113.397
12	election	195	167.482	27	everyone	125	108.455
13	president	289	166.828	28	how	282	103.927
14	help	197	166.607	29	campaign	194	102.224
15	kind	123	166.589	30	Scranton	42	96.987

Top 30 keywords, from the Clinton corpus (FIG 3) (Chen, Yan, & Hu, 2019)

1.4 Sentiment Analysis

Sentiment Analysis is an approach to NLP that identifies the emotional tone behind a body of text (Barney, 2023). It is a widely used approach and as Tetsuaya Nasukawa and Jeonghee Yi it gives enormous opportunities for various applications. (Nasukawa & Yi, 2003) They describe how positive and negative words are chosen from a text, “in the sentence, “XXX beats YYY”, the expression “beats” denotes a positive sentiment toward XXX and a negative sentiment toward YYY.” (Nasukawa & Yi, 2003). This shows the understanding of this work has been around for decades along with NLP itself. This process is much more advanced with

NLTK, it gives words rankings and gives them a negative or positive connotation. It will then add or subtract to the score and if the overall score is over certain number, then it is considered a negative text, or if it stays below the number then it is appositive speech, which is like Nasukawa's example above.

Work done by Prabhsimran Singh, Ravinder Singh Sawhney and Karanjeet Singh Kahlon called Forecasting the 2016 US presidential Elections using sentiment analysis (Singh, Sawhney, & Kahlon, 2017), shows a great use of sentiment analysis. They use data collected from the social network platform, which is twitter, to attempt to predict the future of the presidential election at the time, which was Donald Trump against Hilliary Clinton. They make it clear that it was not the volume of tweets that decided who would win, but more so the sentiment of the tweets themselves. Their results showed that Donald Trump would have more positive and negative sentiment but had a larger net positive score in the end. In their own words they conclude that "The results of our experiments clearly indicate that Donald Trump would be winning the 2016 US Presidential Elections." (Singh, Sawhney, & Kahlon, 2017)

Similarly, a study with similar gains was done on tweets to understand their classes and their homophily (Caetano, Lima, Santos, & Marques-Neto, 2018). In this work, they use sentiment analysis to put them into six classes "We defined six user classes regarding their sentiment towards Donald Trump and Hillary Clinton: whatever, Trump supporter, Hillary supporter, positive, neutral, and negative." (Caetano, Lima, Santos, & Marques-Neto, 2018) and to also get their political homophily into three scenarios. Their results are conclusive, and the sentiment gives us an idea of which class they fall into from the above list. They conclude that those with negative sentiments form a homogeneous community and stay more engaged in the use of common hashtags and retweeting among peers. (Caetano, Lima, Santos, & Marques-Neto, 2018)

Both these works, show how broad a sentiment analysis is. It is very clearly a justifiable process to use; with the election prediction it is spot on as Trump in the end would beat Hillary in the 2026 elections and the work done to understand classes and homophily clearly show that those with negative sentiment would be more active in the negative sentiment community they form or join. There a no major research holes of gaps in the knowledge of sentiment analysis as it is a highly used process in Natural Language processing.

1.5 Craig Zeta Analysis

Craig's Zeta analysis, is a new form to measure textual difference. It is useful to find “marker” words that one author uses consistently, but another author or authors use much less frequency (Hoover, 2010). As Hoover explains, it is a powerful but simple method of measuring differences among authors. It has two sets of data, that are divided into multiple sets that are equally sized. Then it compares both datasets of texts, and it ignored the frequency of the words and look for consistency of appearance. (Hoover, 2010) The Zeta test became popular with computational stylistics to determine authorship of texts, first came from John Burrows, but Hugh Craig's variant has been more influential (Barber, 2021) , which is the reason why it is the Craig Zeta Analysis. With this being the newest of the processing that will be used in this research project, it is also the one with the most gaps in the knowledge area, it has not been used as much as its only slowly starting to enter the mainstream of text analysis. The hope is to understand if this method is a valuable method when it comes to NLP, and to understand it as a process more by using it in my research project. When it came to Craig Zeta Analysis on political speeches, this is where I found a major gap in the area that this report will help change and show the usefulness of Craig Zeta analysis when analysing political speeches. As this is a major gap, this report will be the first to use Craig Zeta analysis to analyse political speeches as it is mainly used in literacy research and will show that it is a useful tool in analysing these speeches and will inspire further researchers to use Craig Zeta analysis.

Chapter 2

Tools and Methods

2.1 VScode and natural Language Toolkit (NLTK)

Arguably one of the most important tools was Visual Studio code as it allowed me to create and experiment with all the analysis tools that I would use. There are multiple reasons why I chose Visual studio code for this project including personal experience and ease of access. From personal experience, I had much experience with VSCode especially as a part of my work placement in Dell technologies which was a part of my 3rd year in UCC which gave me more exposure to the application and consistent use making it a very comfortable space for me to code and making it the right fit for my final year project. Another reason it would suit my final year project is because it makes everything easy to access. The built in terminal allowed me to easily install NLTK with a ‘pip’ command and allowed me to spot any errors

when writing my code. It allowed me to open my file with all my collected speeches alongside my collocation code and all my other analysis tools code for that matter, allowing much faster copying and pasting the data from VSCode straight to a new word file with the analysis data. Both these reasons made the choice easy for this project as the familiarity and ease of access allowed me to effectively analyse all my data for this project.

(<https://code.visualstudio.com/>)

A major part of all my analysis tools was the Natural Language Toolkit (NLTK) Which is described on their website as “a leading platform for building Python programs to work with human language data.” (NLTK, 2023). NLTK is widely used by researchers worldwide, it allows the development of NLP applications that analyse text data (All, 2023). It also says on the NLTK website it has easy to use interfaces for over 50 corpora and lexical resources along with many libraries on tokenisation which breaks unstructured data and natural language text into chunks of information to which particularly was used in this FYP project (NLTK, 2023). It is also an easy install in VScode with a simple command ‘import NLTK’ allowing for the quick use of all the code in the toolkit. The massive choice of all the corpora and lexical resources along with the libraries gave me the choice of how I wanted to analyse my text, so that I was not limited in my choice which led me to choosing collocation, keyword extraction and sentiment analysis to get the data from my chosen speeches.

In terms of overall limitations was learning the code with the natural language toolkit. This took some time, as much time and research was into how to use the code and ensure that it gave me the data I wanted, most of the time was spent experimenting with the code and creating many code bases and having to change them and make them work, until I was able to achieve what I wanted from each of the analyses which was an efficient way to analyse all the data. Some of the things I had to learn to use was tokenisation, creating functions for different elements such as opening the .txt with the speeches so that the code could read it. After much time, writing the code became much easier and allowed for cleaner code and the creation of better functions which lead to better data from my tools.

2.2 Collocation Analysis

Collocation analysis was the first tool of NLP that was identified that could be used for analysing the speeches I have chosen, that would give in depth insights. Collocation as explained in the literature review part of the essay is the frequent co-occurrence of words within a certain distance recognised either side of the specified focal word or node. (Hamed, 2021) As discussed, I had chosen to use VScode as it would allow me to be able to code a

program that would collocate the speeches I had chosen. After gathering all my speeches into a .txt file and opened it into Visual Studio code, the next step was installing NLTK into the code base. This was done by a pip install command into the built-in terminal. After which, I downloaded stopwords which removes words like ‘the’, ‘an’, and ‘I’ which allows for more unique and relevant terms to be identified through analysis. Then I created a way for the .txt file of each speech to be read by the program. After which the main choice for the collocation analysis was to be decided, I had to decide how many N-Gram’s I would use for the project, N-Grams are successive items in a text document that may include words, numbers, symbols, and punctuation (What Is an N-Gram?, n.d). In the end the choice to use 2 N-Grams which is also known as a bigram, which is a sequence of adjacent elements in a string of text. (Bigram vs Trigram: Decoding Common Word Mix-Ups, n.d)

Once the analysis is done and the bigrams are collected, it will then print out the lines like the example we can see below in figure 2.1 below. After receiving the data like this, the next step was to move it from VScode into a new word document so that I could clean the data and present it into this report.

```
Top Bigram Collocations (excluding those occurring only once):  
  
health care  
hope virtue  
even greater  
god bless  
men women  
new age  
new era  
time come  
may new  
us remember  
know america  
us long  
let us  
  
Collocation analysis is complete.
```

Figure 2.1 (Example of collocations from Obama’s Inauguration speech)

Initially I ran into some issues with certain words that would give me no real data from the texts such as ‘the’, ‘and’ and the usual punctuation such as commas. This was dramatically affecting and made the process much longer as I had to comb through much of the initial collocations to find one’s that would be unaffected by these sentence building words. As I

said before, I downloaded stopwords and it made the data much cleaner as it took these words out of the speeches and ensured that all collocations were important words in the text. Another limitation was deciding how many N-Grams and how many of the top collocations I would take from the speeches, because these are small speeches, I first tried to take 4 collocations, but nothing would print, then I tried 3 collocations, but it gave the same result. I was able to only get data on bigrams. While this is limiting to my choice, it also allows me to analyse the few collocations better and be able to focus on the key points in almost a stronger and more focused lens.

2.3 Keyword Extraction

For keyword extraction, the same tools would be used and that is Visual Studio Code and the Natural Language Toolkit (NLTK). At this point, I was familiarised with NLTK coding and the python format that its coded in, which allowed for the tool to be made faster than that of the collocation analysis. With all the speeches in a .txt file from doing the collocation analysis, it allowed me to use the same files for my keyword extraction and sentiment analysis. I started by importing NLTK and stopwords into the file, after which I created the code so that my file could be read by the code which can be seen in Figure 2.2. The file name would go before the .txt. Once the file is read the speeches from .txt will be stored in the variable 'text' which can be called later in the code but is later converted to a new 'word' variable as the text is tokenised which is in Figure 2.3.

```
with open('.txt', 'r') as file:  
    text = file.read()
```

Figure 2.2 (File reader)

```
words = word_tokenize(text.lower())
```

Figure 2.3 (New words variable)

With the text now in the variable 'word' the code filters out non alphabetic words like '123' and stop words from the list of words and calculated the frequency of each unique word in the file. It also covers all the words into lower case, so that it is all formatted the same and allowing for universal data. The code then takes the top 20 words in the text and displays it with the relative frequency of the words, which can be seen below in Figure 2.4.


```
us: 2.01%  
nation: 1.05%  
new: 0.96%  
america: 0.87%
```

Figure 2.4 (Example of Keyword extraction from Obama's Inauguration speech)

Due to the fact this it takes the most used words in texts it made the choice of using stopwords essential, as the sentence building words as described earlier would have been the most used words in the text, so by including this it allowed for more revealing words in a text to be used. The first time that I coded this, it would take the raw frequency of the words within a speech. Meaning that it gave me the overall count of how much a word was said overall in the speech, with the example below in Figure 2.5. If I had used raw frequency, it would have distorted my data as certain speeches are longer than other speeches. As explained in Chapter 1.3 under the heading 'Keyword Extraction', that relative frequency would give greater insight into the keywords as the percentage is analysed rather than the number of times it is spoken in a speech.

1. Nation – 11
2. Us – 11
3. America – 10
4. Freedom – 5

Figure 2.5 (Raw frequency of words in Obama's Inauguration speech)

As stated above, one of the main limitations was my initial use of the raw frequency, which was overcome by using the relative frequency, but this was not the one that lost me much time. Another limitation I ran into was an error in VScode, I had created my code, but it would not run on VScode, after many attempts and much remaking and looking through my code it would not work. After a few days I decided to copy and paste that code into a new python file on VScode and see if that would work, surprisingly the exact same code that would not run worked and it gave me the keywords I was looking for, I had believed it was an issue with the terminal in that python file, but I could never be sure. This was a frustrating moment for the FYP as it slowed down my progress but allowed me to invest time into other areas of my project like the sentiment analysis code.

2.4 Sentiment Analysis

For my sentiment analysis, I was able to bring over 1 of the coding elements from my keyword extraction tool, such as the way of opening and reading the files as seen above in

Figure 2.2. other than that, there was very little that would carry over. I imported 2 libraries which were NLTK and Matplotlib. Matplotlib is a way of visualising data with python, as described on their website it is a library for creating static animated and interactive visualizations in python (Matplotlib: Visualization with Python, n.d). Matplotlib would allow me to analyse my sentiment analysis in a chart and present the data in a more visual way. I then downloaded 'Vader lexicon' into my code which is utilized by the sentiment intensity analyzer which is imported by NLTK sentiment. The 'Vader lexicon' has a pre-build sentiment lexicon, which has sentiment scores indicating whether a word in the lexicons list is positive, negative, or neutral. After such, the code breaks the text into 100-word chunks as seen in Figure 2.6 below. With the words now split into chunks, the code processes the chunk and calculates the overall sentiment score for that chunk. With the sentiment score now received from the text, I then use Matplotlib to create a visualization of all my data with an example of the final data below in Figure 2.7.

```
words = nltk.word_tokenize(text)
chunk_size = 100
textchunks = [words[i:i + chunk_size] for i in range(0, len(words), chunk_size)]
```

Figure 2.6 (Breaking the text into 100 word chunks)

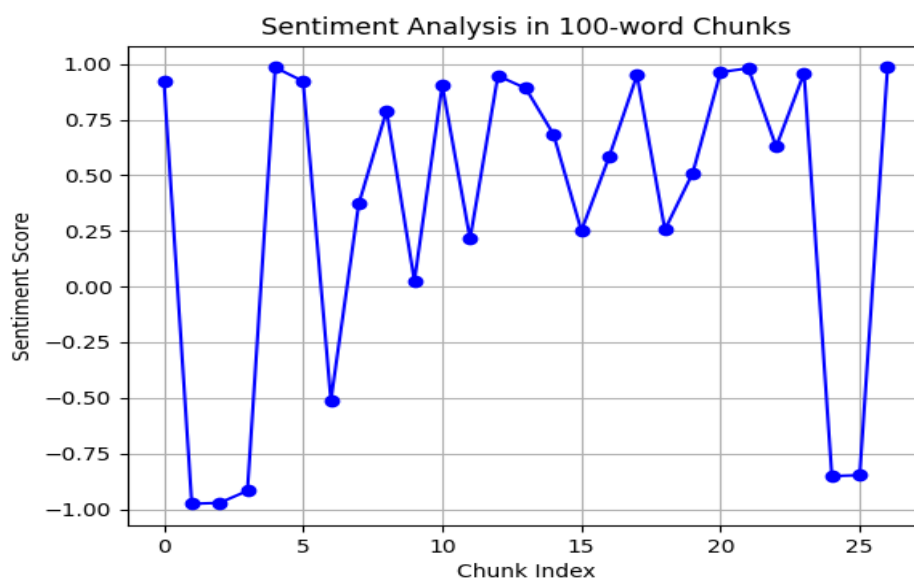


Figure 2.7 (Sentiment analysis on Obama's Inauguration Speech)

Initially I ran into 2 major limitations while doing the sentiment analysis. The first was the analysis itself, initially I created a code base and analysed the entire texts sentiment score but, in the end, this gave me nothing to analyse as they all ended up being positive, which had me

surprised as some speeches tackled scary and tense moments such as John F. Kennedys speech on the build-up of arms on Cuba. This is why I broke the texts into 100-word chunks which gave me great data and allowed me to see how each speech went from negative to positive. The reason I chose 100 over anything else, came from some tests with the numbers. I started with 200-word chunks but there were very few dots on multiple of the speeches, which did not give me great data to work with and then I went to 150 which was better, but still lacked any great insight I could work with. When I landed on 100, it made looking at patterns or trends in the speeches much easier and any number lower gave me to many plots on the graph distorting the data with an example of 50-word chunks below in Figure 2.8 which shows chunks of plots in the centre and at the start. Then this led to the decision of visualisation of the data into a graph which led into my second major issue, which was deciding how to plot this data and how to code it. With Matplotlib it has a vast library for ways to plot my sentiment analysis. I decided on showing a line connecting the markers to give a great way of following the speech and to identify patterns or trends in the sentiment of the speeches. The limitations of my original code were overcome by both these additions of text chunks and plots.

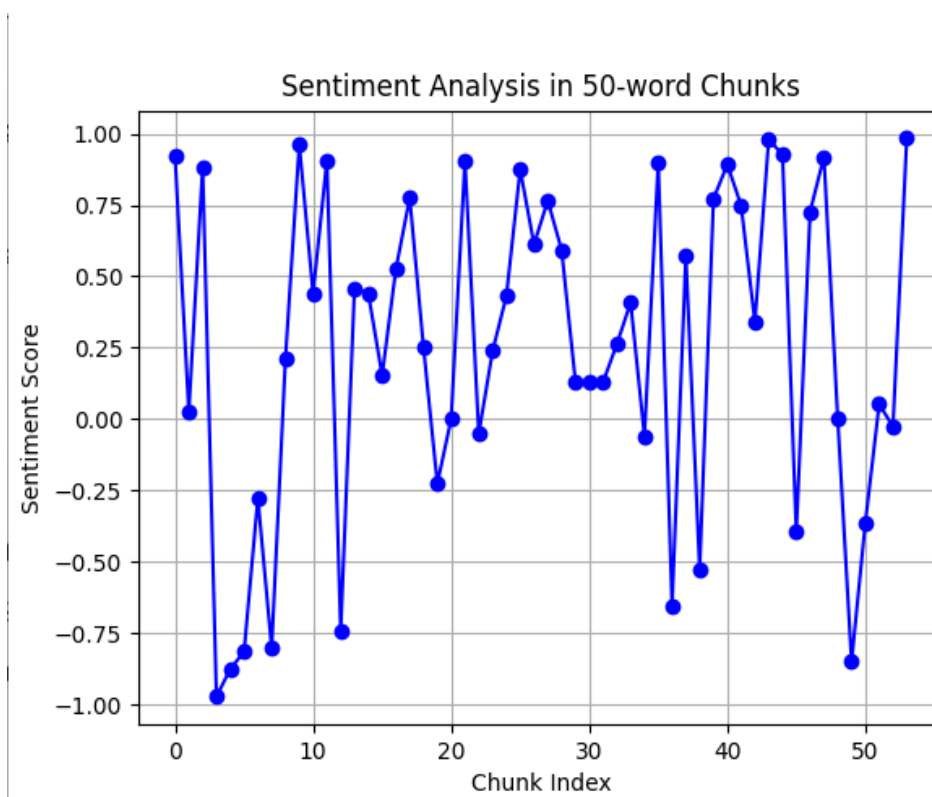


Figure 2.8 (50-word chunk sentiment analysis)

2.5 Craig Zeta Analysis

A tool that I came across in my research was the Craig Zeta analysis, as stated in the literature review this is a tool that has been primarily used in literature analysis and less so in the realm of political science, so that is why I decided to bring this tool into my project and base a research question around it, as I wanted to find out whether this tool can be used for textual analysis in political science. As described here Zeta, is a measure of textual differences that effectively locates an author's characteristic vocabulary 'marker' word that is used consistently by an author and words that are not used at all, but Craig Zeta analysis creates sets of marker words and anti-marker words simultaneously (Hoover, 2010). This updated version is the one chosen for this project.

The first step taken was to create two datasets, a primary set, and a secondary set. I used the same speeches that had been gathered which are the inaugural speeches and the speeches I will compare, which brings the count to 12 speeches. In the primary set we have all speeches from the democratic party presidents, which are Harry S. Truman, John F. Kennedy, and Barack Obama. Meanwhile in the secondary set were the republicans, which are Ronald Reagan, George W. Bush, and Donald Trump. With all the speeches now separated by party the next step was to download R. R as described by their website, is a language for statistical computing and graphics (What is R?, n.d). The reason, R was the choice made for this project is because it is for statistical computing, which is what I was hoping to achieve. It is not something that I had used before, but something I knew would allow me to achieve the goals for this Craig Zeta analysis. The next step was downloading the stylo package. By having R it would allow for easier use of stylo. The stylo package provides an easy implementation of various established analysis in the field of computational stylistics (Eder & Rybicki, Stylo R Package, n.d). With stylo in R, I was to use only 1 function which was the `oppose()` function. This function performs an analysis between 2 given data sets of texts. It generates a list of words significantly preferred by an author, which is our democrats, and another list of words avoided which is the republicans (Eder, Rybicki, Kestemont, & Pielstroem, 2022). I chose this function, as this is exactly what this paper was hoping to analyse. This is why for this paper stylo was used. As, it ensures an easy interface and a way to use the `oppose()` function in R. All these 3 elements allowed me to get the data for this report that would help me understand if Craig Zeta analysis might be able to tell me how both parties differ in terms of their way they prioritise.

There was only 1 limitation that I ran into during this process. When coding, I was not getting any data after using the oppose function. At first, I had believed it was to do with the texts themselves. From the moment I had chosen my speeches I had cleaned them of any texts that was incompatible with VScode, so I had to check to ensure that the texts did not have any that might have been missed throughout. This turned out to not be the issue. After looking at the options that appear once the oppose function is called, is the word slicing option. The option was set to slice at 3000 words, but the issue with this is that it is usually set to slice huge amounts of text, with a much larger word count then what mine was at. After realizing this, the fix to this issue was setting it so that it took the word count into account. In the end the decisions were to set it to calculate on 1000-word slices, once this was changed to 1000 words this issue was fixed and I was able to the results from the crag zeta analysis. With an example of the results below in Figure 2.9.

```
Word preferred:  
  
united  
war  
states  
peace  
only
```

Figure 2.9 (Example of Craig Zeta analysis from democratic presidents.)

Chapter 3

3.1 Analysis of the first research question

In this part of the report, we will investigate the research question “Using Natural Language Processes (NLP), how have United States Inaugural speeches key points changed from cold war presidents to modern day presidents”. I will do this by firstly looking at the collocation analysis, then the keyword extraction and finally sentiment analysis. After such I will conclude the differences between cold war presidents and modern-day presidents to see what inaugural key points have changed. This report will split the cold war presidents together and the modern-day presidents together and analyse them separately and see the differences once all speeches have been analysed.

3.1.1 Collocation Analysis

1. decent satisfying	1. beyond doubt	1. hopes goals
2. north atlantic	2. bear burden	2. martin treptow
3. achieve decent	3. begin anew	3. men women
4. armed attack	4. forms human	4. fate fall
5. courses action	5. go forth	5. happiness liberty
6. related agencies	6. finished first	6. provide opportunity
7. standards living	7. vice president	7. dreams hopes
8. determined work	8. call bear	8. special interest
9. false philosophy	9. first one	9. fellow citizens
10. four major	10. generation americans	10. tax burden
11. greater production	11. fellow citizens	11. act worthy
12. major courses	12. ask country	12. vice president
13. technical knowledge	13. shall always	13. go away
14. democratic principles	14. let every	14. believe fate
15. desire determined	15. us begin	15. children children

Figure 3.1 (Truman's Inaguration speech), 3.2 (John F. Kennedy's Inagurational speech) and figure3.3(Ronald Reagan's Inaguration Speech)

Firstly I will look at Truman's collocations in Figure 3.1 above on the far left side. The major key point that I can see from these collocations is based around the protection of American democracy and the advancement in the human race. In terms of protection the backdrop of the past few years for Truman with the second world war, would leave scares and the use of the phrase "north atlantic" would indeed be in reference to the new alliance of western nation's to curve Soviet aggression. This is why words like "democratic principles" appear in the same list as these are the values that need to be secured for Americans. We also see the key points around advancement take shape with collocations like "technical knowledge" and "greater production" reference the drastic change that America is facing, with it being a global superpower with a great production of consumer goods and the technical knowlegde which has grown with major breakthroughs in science like the atomic bomb. This overarching key point of security feed into Trumans policies abroad and the ideolgal war the United States are fighting againt as Truman says a "false philosophy" of communism which we see as one of the collocates in his list.

Next, I will look at Kennedys collocation's in Figure 3.2 above in the middle. Straight away one can take away is that the key points of this speech is about the people and the struggle the United States is facing at this point in time. Collocates such as "fellow citizens" and "generation americans" is all about relating to the people. His collocates are about the burdens they carry when he says "bear burden" which clearly shows he understands the

people have burdens and that the populace is carrying a weight of the cold war on their shoulders. Also quite interestingly in his collocates is a call for action, the use of the words “go forth”, “ask country” and “us begin” show a simple key point of now is the time of action, of going out there and doing the work for the country to lift these burdens as people of the United States and this is there time to begin this work. By looking at these collocates it seems that JFK is a dealer in hope and action for his people as he calls upon the people to act.

Lastly I will look at Ronald Reagan’s collocation above in Figure 3.3 on the far right side. Much like that of John F.kennedy we see Reagan’s key points around hope for the people, but he does it in a different way to JFK. At this point for the United States the economy has slowed but the fear of war has also diminished for many people, but the economy slowing down lead to a great deal of pain for many. We see him mention hope twice in his speech with the collocates “hopes goals” and “dreams hopes”, this is not a mistake for “hopes” to be used especially when tied to “goals” and “dreams” which carry a positive meaning when paired with hope. We also see what is another huge key point in the collocate of “tax burden” which shows Reagans want for change of the economy and to lower the “tax burden” for the people in his country which in turn would hopefully help the American economy. We see these two major key points of hope and change for the people take center stage from these collocations list.

1. rides whirlwind	1. health care	1. men women
2. whirlwind directs	2. hope virtue	2. god bless
3. god bless	3. even greater	3. united states
4. president clinton	4. god bless	4. transferring power
5. public interest	5. men women	5. stops right
6. common good	6. new age	6. bring back
7. america best	7. new era	7. every city
8. service nation	8. time come	8. nations capital
9. america never	9. may new	9. president obama
10. citizens every	10. us remember	10. together make
11. nation promise	11. know america	11. american workers
12. story story	12. us long	12. states america
	13. let us	13. make america

Figure 3.4 (Goerge W. Bush’s Inaguration speech), 3.5 (Barack Obama’s Inagurational speech) and Figure 3.6(Donald Trump’s Inaguration Speech)

For the modern day presidential speeches, we will look firstly at Bush's speech in Figure 3.4 above on far left. There are many interesting collocates in Bush's list, often then the few that are essential when taking on the speech such as mentioning "bill clinton" who was the President before George W. Bush came into office. The interesting ones to me are "rides whirlwind" and "whirlwind directs" which is an indication that a key point of this speech is the fact that America can ride the storm and be able to direct that storm indicating that they have the power in their hands to survive the turbulent years that have come and gone and that are approaching. Much to the same mark we can see a key point of creating a narrative for the American people with collocates like "service nation", "america best" and "story story", this is all about creating a narrative, that the American people can follow and understand, that the nation is a nation of service and is the best in the world. These 2 key points really resonate with the American people, that they control the "whirlwind" and can "direct" it to be better for them, and that this narrative is one they can control. Showing that these big key points that he underlines in his speech.

Next, I will look at Obama's collocation in Figure 3.5 above in the middle. With the Great Recession and the war in afghanistan one can understand the main key point that we can see in Obama's inagural speech which is about a "new age" and "new era". With the issues of the past and America still in econmic decline, it comes with no wonder that America needs hope for the future and for this administration. Obama provides this in indicating that this is a new "era" and "age", closing the window on the past and opening a new door for the future so to speak. It is no mistake that the word "new" appears 3 times in this list with the additional collocate of "may new". This is a choice made by his team or by Obama himself, to try bring hope to the people for a new and brighter future. Another key point in his speech is "health care" for Obama healthcare was an important electoral point which he promised to bring changes, much like when he said "President Obama promised that he would make quality, affordable health care not a privilege, but a right." (Improving Health for All Americans, nd). With this we can understand why it would be placed first in the list of collocates as its an important key point of his speeches.

Laslty I will look at Trump's collocation above in Figure 3.6 on the far right side. One of Trumps major election promises was putting America first and to in his words make America great again. We can see this key point shine through by looking at his collocates. We see references to "America" and the "States" in 4 collocates with collocates like "united states", "american workers", "states america" and "make america". These indicate his position when

it comes to America, he wants to put “american workers” first, and sees them as a very important part of America and that his key points are based around the people of america and not a wider international population. Apart of this key point about America in his eyes, is a lot of reminicing about America and the nations past and we see this from “bring back” and “make america”, these clealry show that Trump does believe the nation has lost there power, which to me is a great turn back from Obama and this new era he was pushing and then to have Donald Trump to be saying they need to go back to the way things were and to “make america” great again.

Overall, the collocates reflect their unique style and focuses, that highlight continuity with the past administration but leaving space for their own visions of change. There key points range massivly but they do indeed have some simliarites. For contrast there are some outliers here, in Donald Trumps keypoints in his speech it is all about looking to the past meanwhile the other presidents look to a new future. Besides that we see an outlier of Obama with his key point on health care, which does not appear elsewhere in any of the collocates and finally in Trumans speech which has a major focus on the protection of democracy and advancement and security of freedom and democracy and finally Bush, who is about creating a narrative for the american people, this change in cold war to modern day presidents is signifigant. These outliers paint a picture, that the way speeches are made has changed, modern speeches focus on different smaller points instead of a what can be seen as the bigger picture which is hope for the future, Truman talks about democracy and the need to protect it, Trump is about going back to the past and reminising, Obama is about healthcare and Bush is about the narrative of America past and future which shows the unique vision is more important in modern day speeches. I say this as we see one major comparision between cold war and modern day presidents and that is hope. In JFK speech we see hope and action as his key point, in Ronald Reagan’s we see hope and economic focus and his key point and finally we see Obama about hope and change. Two cold war presidents focus on hope meanwhile only one modern day president does. There can be two explanations back in the cold war presidents needed to present a more hopeful image to calm the nation more then modern day presidents or that modern day presidents are more about vision and their goals and less focus on giving hope to the people of America. These chosen key points in each speech, highlighted through collocates, are like strategic arrows aimed at getting their messages right into the hearts and minds of the American people, some focus on smaller aspects meanwhile some present an idea of hope for the nation.

3.1.2 Keyword Extraction

1. nations: 1.99%	1. let: 2.34%	1. government: 1.51%
2. peace: 1.18%	2. us: 1.75%	2. must: 0.89%
3. people: 1.09%	3. world: 1.17%	3. believe: 0.89%
4. freedom: 1.09%	4. sides: 1.17%	4. world: 0.80%
5. free: 1.00%	5. new: 1.02%	5. people: 0.80%
6. united: 0.90%	6. pledge: 1.02%	6. americans: 0.80%
7. peoples: 0.81%	7. citizens: 0.73%	7. one: 0.71%
8. democracy: 0.81%	8. power: 0.73%	8. time: 0.71%
9. must: 0.81%	9. shall: 0.73%	9. freedom: 0.71%
10. new: 0.72%	10. free: 0.73%	10. work: 0.62%
11. believe: 0.72%	11. nations: 0.73%	11. today: 0.53%
12. security: 0.72%	12. ask: 0.73%	12. nation: 0.53%
13. countries: 0.72%	13. president: 0.58%	13. upon: 0.53%
14. program: 0.63%	14. fellow: 0.58%	14. let: 0.53%
15. justice: 0.54%	15. freedom: 0.58%	15. man: 0.53%
16. help: 0.54%	16. man: 0.58%	16. america: 0.53%
17. men: 0.54%	17. first: 0.58%	17. president: 0.44%
18. communism: 0.54%	18. americans: 0.58%	18. called: 0.44%
19. economic: 0.54%	19. war: 0.58%	19. national: 0.44%
20. resources: 0.54%	20. peace: 0.58%	20. children: 0.44%

Figure 3.7 (Truman's Inaguration speech), 3.8(John F. Kennedy's Inagurational speech) and figure3.9 (Ronald Reagans Inaguration Speech)

The first president keywords we will look at and examine for key points is Harry S. Truman's speech above in Figure 3.7. From these keywords we can understand what his inagurational speech was about. In the collocation section we spoke about his collocates indicated towards key points of protection of democracy and this is reinforced by the keywords extracted from the speech including "freedom", "democracy" and "security". These words are keywords as they are important and overall they take up 2.62% of the total word count of the speeches. Alongside this we see another impotant part of this. He also clearly uses keywords like "nation", "peoples", "free" and "countries", this is to indicate the very ideas that get the US involved in affairs abroad to prevent the spread of "communism" which appears in the keywords list above. It perfectly ties into the collocates keywords of protecting democracy but builds on it to not just include at home but to incinutate protecting abroad and protecting the "free" "peoples".

In John F. Kennedy's inaugural speech above in Figure 3.8, we can see some very interesting key words, it follows much of the same lines as Truman's key words with words like "freedom", "peace" and "nation" which again appears highly in the list. It does differ and seem to have an approach of being relatable to the people and calling for action much like in the collocations with words like "people", "common" and right at the top "us". These words relate to the people of the United States and by using "us" gives the struggle they will face a unifying front that includes the president. We can clearly see that it is much like Truman's who also uses "us" in a top spot of the keyword's extracted. As stated we see an emphasis on action alongside this we see a call for change, which can be collected from keywords like "new" and "work" with incuinating a "new" era or "generation" that must "work". We see his keyword of action and hope still evident in the keywords but with the addition of change, and also with a more unifying approach to the issues ahead.

With Ronald Reagan's keywords extracted from Figure 3.9 we see a very different approach from his key points. From the offset we can see a very distinctive approach, but we do see some similarities the usual term "nation" and "freedom" which clearly is a staple of cold war speeches from this time, but we see a big change in the first word "government", this is a major part of his belief that there should be no big government which we shall look more into in Chapter 5 alongside the use of the word "children", which seems to be an approach to relate to the populace. This word indicates a key point of his around "government", but in collocation it was around economic focus and hope which is hard to get from these keywords, but much like JFK we see what seems like a call for the country to act for change with reference to "must", "word", "let" and "national" with that distinct unifying approach almost missing with no use of the word "us" in the top 20. We also see a shift in theory, there is less mentions of the world beyond and of "war" in JFK speech and "peace" in both prior speeches, instead there is only internal key points mentioned, the biggest keypoint change in cold war speeches is this shift from outside affairs in inaugural speeches, it is as if, that the longer the cold war wained on, that this kind of keypoint around the cold war and issues abroad became less popular, we see it is very evident with Truman and see a drop in this point in JFK but in Reagan's speech it is completely gone and replaced by other words for example "government".

Now I will look at the modern day presidents and start firstly by looking at George W. Bush's speech below in Figure 3.10. We see once again the word "nation" and "us" take centre stage in his speech, then we can notice that the word "america" is one of the top words extracted,

this is a great leap in this word from both JFK and Truman and even higher than Reagans. This is a time where the Soviets are gone and it is a golden age of moderate peace for the world, leaving room for this “american” keyword as policies can shift to domestic issues. Alongside what this report said about Bush’s collocates referencing of his building key point around a narrative, is reinforced here by the keywords with the word “story” ranking so highly in his list at 4.

Next we look at the keywords extracted from Barack Obama’s speech in Figure 3.11 below. From the collocates from Obama’s speech, we could concur that Obama was driving home the keypoint that his administration would bring something new for the people, once again this point is backed up from the list below, we see the word “new” appear in the 3rd place of the list, showing how important his key word is for the Obama inaugural speech, he is clearly trying to get his point of “new” across to the people, to emphasize that this is a new administration that could bring a new change to the United States. Unusually we see the first mention of “peace” in Obama’s speech, which is a word we see used in 2 out of the 3 cold war speeches, this can be attributed to the war in Afghanistan and it shows, that when America is at war, or even in an ideological war like the cold war, the word “peace” goes interchangeably in the inauguration speech. Besides this, we see that, the collocates “health care” which are an important part of his administration does not present itself as a keyword, which shows it is used together a lot in his speech and not as single words.

With Donald Trump his first keywords indicate a lot about his key points in his speech. From the keywords extracted in Figure 3.12, words like “america”, “american” and “americans” take up 4.8% almost 5%, this is a high use of these keywords, that much like his collocations show that his key points are based around the people, as I said it’s almost like an American first policy all most, he wants to relate to the people and to the Americans to appeal to them and he did this a lot during his campaign but has carried it on into administration. Also noted from his collocations, was the fact he wanted to return America back to its golden days with phrases like “make America great again” being a huge staple in his election and now we see it in the keywords. We see this in one word “back”, he wants to present to the people a feeling of nostalgia of the past and prior American dominance which was draining at this point. Overall, his key points focus on the people, to relate and to get them feeling nostalgia of the past.

1. nation: 1.42%	1. us: 2.01%	1. america: 2.57%
2. us: 1.42%	2. nation: 1.05%	2. american: 1.71%
3. america: 1.30%	3. new: 0.96%	3. people: 1.28%
4. story: 1.30%	4. america: 0.87%	4. country: 1.28%
5. citizens: 1.17%	5. every: 0.70%	5. one: 1.14%
6. country: 1.17%	6. must: 0.70%	6. every: 1.00%
7. must: 0.78%	7. today: 0.61%	7. world: 0.86%
8. every: 0.78%	8. people: 0.61%	8. great: 0.86%
9. common: 0.65%	9. less: 0.61%	9. back: 0.86%
10. new: 0.65%	10. world: 0.61%	10. nation: 0.86%
11. many: 0.65%	11. let: 0.61%	11. never: 0.86%
12. freedom: 0.65%	12. time: 0.52%	12. new: 0.86%
13. promise: 0.65%	13. work: 0.52%	13. president: 0.71%
14. americans: 0.65%	14. common: 0.52%	14. nations: 0.71%
15. know: 0.65%	15. generation: 0.44%	15. protected: 0.71%
16. never: 0.65%	16. day: 0.44%	16. across: 0.71%
17. courage: 0.65%	17. know: 0.44%	17. right: 0.71%
18. yet: 0.52%	18. spirit: 0.44%	18. dreams: 0.71%
19. power: 0.52%	19. words: 0.35%	19. make: 0.71%
20. american: 0.52%	20. peace: 0.35%	20. americans: 0.57%

Figure 3.10 (Bush's Inaguration speech), 3.11 (Barack Obama Inagurational speech) and figure 3.12(Donald Trumps Inaguration Speech)

In the end, by looking at these keywords extracted from all of the speeches we can see some change in the keypoints from cold war presidents to modern day presidents. Firstly we can see an interesting growth in what can be considered more national tendencies, we can deduce this from the growth of certain keywords from the lists above and that is "america", "americans" and "american". From first glance one can see why it might be important in an American presidents speech but what is interesting is that this word does not feature on the top 20 in Trumans, in JFK's speech one of the words rank at 18, then in Reagan's it rises to be placed at 6 and 16, but when we enter modern day presidents we see a major jump, in Bush's it all 3 of those words appear. Placing at 3, 14 and 20, in Obama's we see this only place once at 4 and in Donald Trump's speech its placed at 1st, 2nd and then end at 20. This is a slow and steady rise of these words and I do not believe it is a mistake. It is a growing key point for United States presidents all most as if the wider America spread their influence around the world, they think the focus is less on America's home issues. I see this an important key point that has changed, in the cold war they focused more outwards meanwhile

in modern day politics it seems they focus more on the people within their own borders. Another key point that has changed is based on unity. In cold war presidential speeches, mostly Trumans and JFK's, it seems they do not only focus on themselves but mainly the wider community, as if to unite them in a common goal, which was to outlast communism, this changes slightly in Reagans speech but the use of keywords like "world" show that this empahsis is still there. Meanwhile modern day do not seem to carry the same sentiment, and seems to key it more internal and want to unify the American people more, with words like "us" having a strong presence in these speeches but also appearing in cold war, but no reference to the "world" appearing as much as it did in the cold war speeches. In conclusion, the key points have shifted from Cold War presidents to modern-day presidents' inaugural speeches in the way that they are becoming more internally focused. Today's leaders not only emphasize the needs of the American people but also center their messages on unity among the populace rather than on global concerns.

3.1.3 Sentiment Analysis

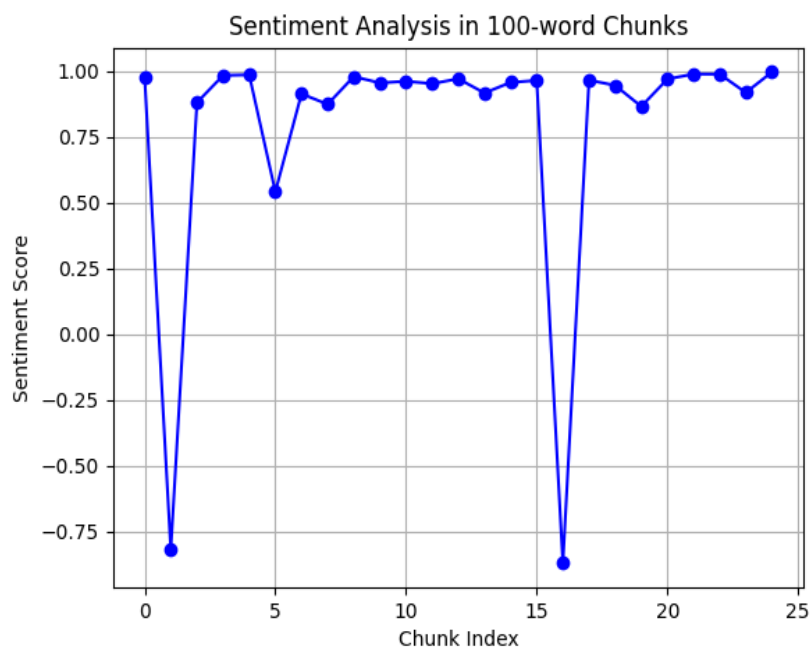


Figure 3.13 (Harry S. Truman January 20, 1949: Inaugural Address)

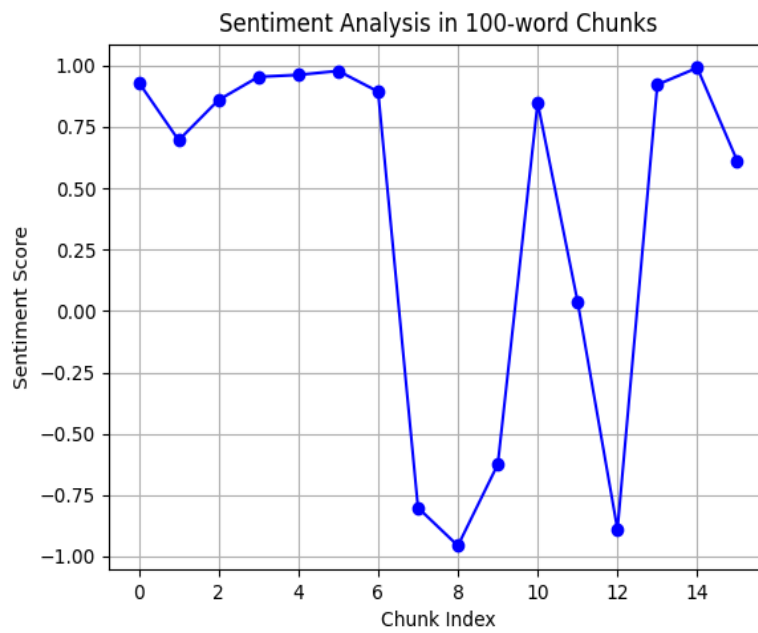


Figure 3.14 (John F. Kennedy January 20, 1961: Inaugural Address)

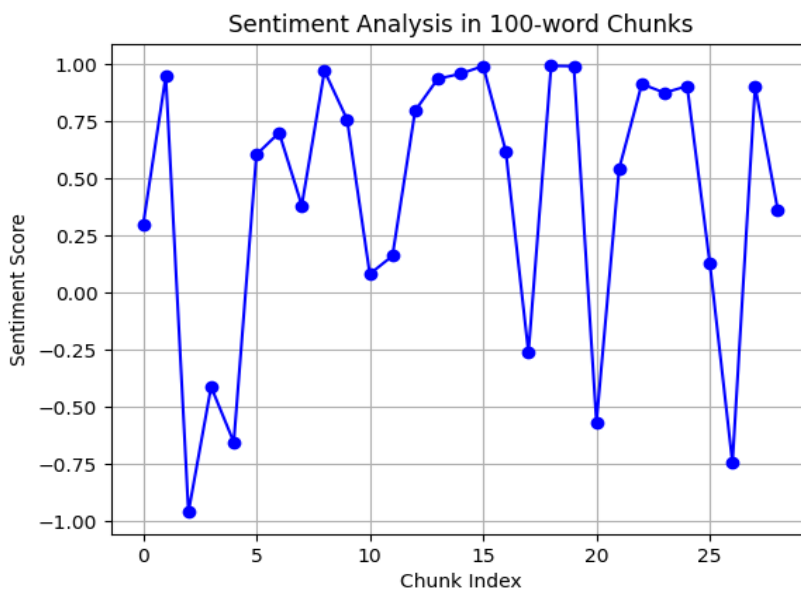


Figure 3.15 (Ronald Reagan January 20, 1981: First Inaugural Address)

Now let's look at Cold war sentiment analysis in their speeches, with Truman in Figure 3.13, JFK in Figure 3.14, and Reagan in Figure 3.15. The cold war speeches have an average of 23 100 word in their speeches. Both Reagans and JFK's speeches drop to a -1.00-negativity score with Truman staying at a -0.75. We also can see that Truman's speech stays high with sentiment as most of his chunks stay close to a 1 score, meanwhile JFK's and Reagan's jump

around in sentiment with Reagan's speech staying around the negative line, in many of his chunks in the middle, with JFK'S being the most negative due to the length of the speech with 4 staying below the negativity line and 2 staying as the lowest score of -1.00. Besides the negativity we can see some patterns from these speeches, they appear to all drop in negative very soon, as they tackle their issues, with another drop around the $\frac{3}{4}$ of the way in and finally finishing with a positive sentiment to inspire hope.

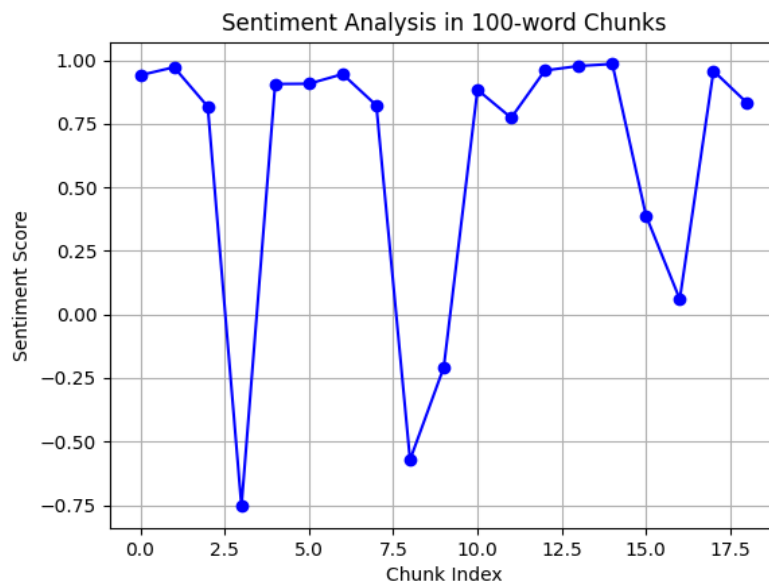


Figure 3.16 (George W. Bush January 20, 2001: First Inaugural Address)

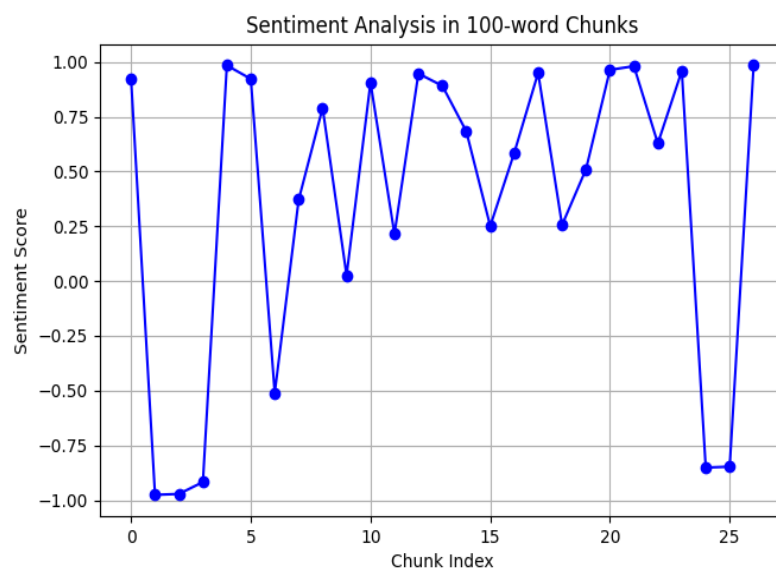


Figure 3.17 (Barack Obama January 20, 2009: Inaugural Address)

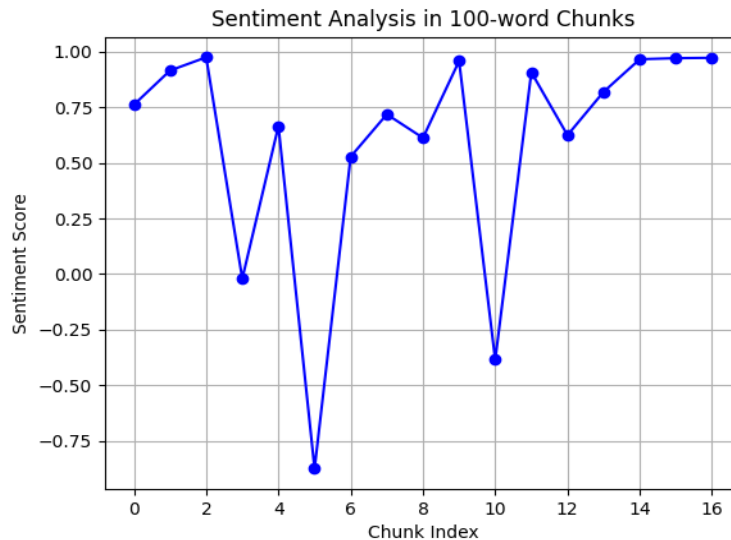


Figure 3.18 (Donald Trump January 20, 2017: Inaugural Address)

In examining the speeches from Bush in Figure 3.16, Obama in Figure 3.17, and Trump in Figure 3.18 above, we find that modern-day speeches, on average, consist of 21 chunks of 100 words each. This count is lower than during the Cold War era. Notably, only Trump's speech shows a negative sentiment score of -1.00, while all three speeches hit a sentiment score of 1 at some point, indicating a more positive tone compared to Cold War speeches. Trump's speech stands out as the most negative, with Bush's and Obama's speeches leaning more positively. However, Obama's speech follows an interesting pattern of starting positively and then shifting towards a more neutral score. Like Cold War speeches, these modern speeches begin positively, dip into negativity, and then end positively again.

Overall, we see that modern day speeches are shorter with an average of 21 against the cold war average of 23, we also see that 2 speeches from the cold war drop to -1.00 sentiment, compared to modern days, 1 speech that drops that low. When it comes to patterns, we see that all the speeches seem to all fall to a negative around the start, and all end a positive sentiment note.

3.2 Conclusion

Now we need to examine the overarching themes. From the patterns in their language use, it becomes clear that most presidents aim to inspire, provide hope, and outline their major policies for their administrations. However, this has not prevented some changes from taking place. The significant shift we observe in collocations from the Cold War era to modern-day

presidents indicates a change over time. Initially, there was a focus on the broader global context, as seen in Truman and JFK's language choices. Yet, as time progressed, this focus seemed to narrow. Bush emphasized narrative, Obama was focused on healthcare, and Trump often looked back to the past. This shift is also apparent in the keywords extracted from their speeches. Over time, there has been a noticeable change in the keywords used. Notably, since Truman's inaugural speech, there has been a rise in keywords related to the American people. This shift suggests that America is becoming more internally focused, with a greater emphasis on domestic policies and home defence, as evidenced by the US moving away from global rhetoric and concentrating more on internal affairs. We also see in the sentiment analysis, that cold war speeches appear to be longer and more negative, which can indicate that the issues that were facing the United States were a bigger threat than what the modern-day presidents have to face, or that the sentiment has become more positive as the years have gone by.

Chapter 4

4.1 Analysis of the second research question

In this part of the report, I will answer the research question of “How do Cold War presidents and modern-day presidents tackle shared global issues or internal issues in their speeches?”. I will do this by firstly looking at the collocation analysis, then the keyword extraction and finally sentiment analysis. Firstly, I will compare Harry S. Truman “Truman Doctrine” speech to George Bush’s remarks on freedom in Iraq and the Middle East speech, then John F. Kennedy Speech on the build-up of arms in Cuba to Barack Obama speech to the people of Cuba. Finally, Ronald Reagan speech to the nation on the campaign against drug abuse to Donald Trump’s remarks on combating the Opioid Crisis. At that I shall compare the differences between both cold war presidents and modern-day presidents to see how they handle internal and global problems.

4.1.1 Collocation Analysis

Let’s firstly look at Harry S. Truman’s speech below in Figure 4.1. In this speech, he discusses sending aid to countries like Greece and Turkey that are both in the midst of a communist civil war, and to do that he must convince congress and this is evident with collocations like “ask congress” and “congress authorize”, besides this we get that strong sense that there is a major increase in tensions and American “foreign policy”, the aim is to restore the “status quo” which has already been rocked since the second world war, it is more

evident the plan is to create a new status quo. He presents it to ensure “national integrity” by protecting freedom. The goal is not restoring the economy but also to send “supplies equipment” to these nations to fight back against these communist groups. Interestingly the mention of the “middle east”, this shows how important even in 1947 this region of the world is, even though it is not very developed now. It gives a sense of urgency to this matter, and we see his goals as protecting the American way of life abroad which can be seen from the collocate “way life”. In the end a picture of how important working together with their allies abroad is to the Americans and at this point, they want to ensure the integrity of their allies abroad and stay intact from communist “totalitarian regimes”.

In George W. Bush’s speech below in Figure 4.2 which is based around American circumstances in the middle east. We see this region is still a focal point of American foreign policy more so now than ever, with the rise of terror groups and the strong oil reserves of the countries in the region. We get an indicator as to the reasons why America is focusing on this the Middle east with collocates like “prison camps” and “oppressed peoples”, we see why they believe their cause is just. In 2001, 911 was followed by the invasion of Afghanistan, which by this speech is a pacified country for now at least and in March of 2003 America was involved in conflict in Iraq. We see him almost justifying their invasion with these collocates. The collocate “soviet union” is interesting, at this point the USSR is dissolved and not a threat no more, and its place in this speech is almost uncharacteristic due to them having no hold on global relation, one could argue it is a way of comparing the lives of those in the middle east to that of the Soviet Union. His collocates show us that America foreign policy, is more aggressive at this time, and that America must be protected by action abroad.

1. improving public	1. turning point
2. public administration	2. united states
3. status quo	3. three decades
4. supplies equipment	4. willingness sacrifice
5. middle east	5. years ago
6. totalitarian regimes	6. prison camps
7. foreign policy	7. oppressed peoples
8. reports american	8. europe asia
9. great britain	9. historians future
10. life based	10. may god
11. national integrity	11. working together
12. way life	12. ronald reagan
13. able provide	13. rule law
14. ask congress	14. soviet union
15. congress authorize	15. way toward

Figure 4.1 (March 12, 1947: Truman Doctrine) and figure 4.2 (November 6, 2003: Remarks on Freedom in Iraq and Middle East)

These similar global issues are tackled in the president's own style and show a change in American foreign policy. First Truman wants to protect America by playing a supporting role abroad with aid and supplies, but Bush's policy is one of action, Bush clearly believes his wars are just and aim to free the people of the Middle East.

Next, we will look at John F. Kennedy's speech below in Figure 4.3. This speech is based around the growing tension with the movement of Soviet Union nuclear missiles to Cuba which is why we see words like "military equipment", "missile sites" and "capable striking". Evidently, we see a resolve in his collocates with "resolution congress" and "call upon", he tackles this issue by showing the dangers of this course of action by the Soviets, he wants the people to understand the magnitude of this by really giving the soviets a scary face with words like "foreign domination". Interestingly there is no mention of Cuba in any way, so we see him facing more on the threat and not the players in the crisis, but noticeably he does mention it as a region with "latin america" which includes all of south America in this including Cuba. He clearly wants people to understand this threat and does not try to downplay the risks at stake for the people.

In Barack Obama's speech in Figure 4.4 we see a very different approach. America and Cuba's relation were always hurt by this crisis and had never fully recovered since the "cold war" which is mentioned. It is also important to understand how close they are, "miami havana" get mentioned, due to the fact that these cities are the closet to each other with only

330 miles in between them. These nations are close, and we get a great sense of the reason why Obama visited this country. He tackles the issue by trying to return relations as he says “normalize relations” and “normalizing relations”, this goal is a major change in American policy, he goes to restore these relations and to try and gain a working relationship with them. Aside from this the collocate “health care” only reinforces how important this is for Barack Obama, as this even appears in this speech to Cuba, only showing this key point stands and is his goal.

Overall, we see another change in America's foreign policy and how they tackle this global issue. JFK presents it as a threat and at the time it was but left very little room for reproachment meanwhile for Obama reproachment and normalizing relations is the only goal.

1. status quo	1. kind cooperation
2. fellow citizens	2. muchas gracias
3. impose system	3. everybody agrees
4. resolution congress	4. paid tribute
5. capable carrying	5. pueblo cubano
6. capable striking	6. cold war
7. foreign domination	7. health care
8. first latin	8. normalize relations
9. missile sites	9. normalizing relations
10. station strategic	10. year father
11. call upon	11. children grandchildren
12. latin american	12. start small
13. military equipment	13. taking place
14. system upon	14. finlays work
15. united states	15. miami havana

Figure 4.3 (October 22, 1962: Address on the Buildup of Arms in Cuba) and figure 4.4 (March 22, 2016: Remarks to the People of Cuba)

Next, we will look at Ronald Reagan's speech in Figure 4.5 below, which is based upon the growing drug crisis in America. Reagan had a strong campaign against drug use in America, he continued it from Nixon, but expanded the reach of the war and focused on criminal punishment (War on Drugs, 2024). This strong man policy does not appear to focus on this speech, but instead a very opposite tone is strived upon the American people. We see it tied to “foreign countries” as a lot of drugs were being smuggled in, but this is the one mention of it, besides this, he seems to wish to paint this issue as one that can be overcome, not through arrests or crackdown, but through a “national crusade” which is a very interesting use of the

word, as it paints it as an issue that the whole country has a role in playing to overcome this crisis, he drives this home with words like “fight drug” and “working together”. There is that strong sense of nationalism in this, to put the nation to work in helping themselves to overcome this issue.

When it comes to Donald Trump’s speech in Figure 4.6 on the opioid crisis in which America was facing, we see an utterly different approach. In his speech, there is a lot of words that do not seem to tie to any main point in this speech like “blue ribbon” and “Jim Jean”, these do not add to any key points that he makes throughout, but we do see where he seems to want to focus. With collocates like “gang members” and “southern border” there is this sense that, he wants to put the blame on these elements and not focus on the problem as a crisis that can be overcome by individual betterment and community betterment. He does seem to focus on “second chances”, which would be in reference to stopping drug use, but this is slight mention of what he wishes for the people, and instead looks abroad and at gangs for blame.

1. foreign countries	1. blue ribbon
2. government private	2. jim jean
3. law enforcement	3. little bit
4. men women	4. short period
5. white house	5. adapt pharma
6. working together	6. already saved
7. president lincoln	7. gang members
8. last years	8. god bless
9. young people	9. knows better
10. national crusade	10. lawrence massachusetts
11. alcohol abuse	11. loved ones
12. drug criminals	12. second chance
13. fight drug	13. southern border
14. something special	14. tie wall
15. gave us	15. federal level

Figure 4.5 (September 14, 1986: Speech to the Nation on the Campaign Against Drug Abuse) and figure 4.6 (March 19, 2018: Remarks on Combating the Opioid Crisis)

Both Reagan and Trump tackle this internal issue in different ways. Reagan wants to paint it as an issue for the people, a crisis that can be overcome by everyone which is why he paints it as a “national crusade”, meanwhile Trump does not take this level of high ground, instead he looks at members of gangs and points blame to the “southern border”, he does not seem to try and unite the American people against the issue unlike Reagan who succeeds in this task.

When looking at the difference in Cold war presidents and Modern-day presidents and how they tackle these global and internal issues we can see different approaches. Firstly, this sense from cold war Presidents of unity still stands here, Truman advocates Unity of nations to dispel communism and Reagan wants Unity of people to fight the drug crisis in their country. The only outlier being JFK who focuses more on informing the Americans of the risk, he uses his words to ensure support and to present a strong front to Cuba and the Soviets. Meanwhile, modern day presidential speeches do not take on this unifying approach unlike Obama who seems to turn that hard front on Cuba made by JFK into a path of normalizing relations and presenting that front of Unity between nations. Meanwhile with Trump and Bush, it is evident they want to justify and blame, Bush tries to justify his wars abroad as he is freeing “oppressed peoples” meanwhile Trump blames “gang members” and the “southern border” for the drug crisis in his collocates. It is clear from the collocates that each presidents have a different approach to their global and internal shared issues.

4.1.2 Keyword Extraction

When it comes to the keywords extracted from their speeches, we will start with Harry S. Truman speech in Figure 4.7. From looking at these keywords, we see certain words we would expect in this speech “greece”, “greek” and “turkey”, at these give us the context of the affected regions. We still get the same hand that we do from the collocates, it is a very unifying speech with words like “united”, “peoples” and “nations” all being stated, but from the keywords we get a stronger sense of the aims of the speech. As stated by the collocates, it is clear that an important goal of this speech it to provide “assistance” to these nations and protect them, but we get a sense of exactly what this goal is and that is to provide “economic” and “financial” “aid” to these countries, we see that this is the goal, it is less on giving them weapons and more on giving them the necessary assistance to protect themselves as the greatest defence of state is proving it can build and work.

When it comes to Bush’s speech in Figure 4.8, we see a much different story from his collocates. We see a lot of collocates based upon the principles of the United States such as “freedom”, “democracy”, “liberty”, and “democratic” all over this speech, as stated in the collocations, it seems he wants to focus on justifying American actions abroad in the Middle East, and these words give me an indication that this is his goal. These words mirror what the US is and what they want to spread into these countries. Words like “free” “women”, join this point and emphasis what he does believe to be the goals of American troops abroad.

1. greece: 3.09%	1. freedom: 1.86%
2. united: 1.94%	2. democracy: 1.60%
3. government: 1.59%	3. middle: 1.28%
4. greek: 1.50%	4. people: 1.06%
5. turkey: 1.41%	5. east: 1.01%
6. states: 1.33%	6. nations: 1.01%
7. assistance: 1.24%	7. world: 0.96%
8. free: 1.24%	8. liberty: 0.96%
9. world: 1.06%	9. national: 0.74%
10. upon: 1.06%	10. free: 0.69%
11. must: 0.97%	11. women: 0.69%
12. nations: 0.97%	12. political: 0.59%
13. economic: 0.88%	13. every: 0.59%
14. congress: 0.80%	14. rights: 0.53%
15. financial: 0.80%	15. democratic: 0.53%
16. aid: 0.80%	16. many: 0.53%
17. peoples: 0.80%	17. time: 0.48%
18. one: 0.62%	18. yet: 0.48%
19. countries: 0.62%	19. nation: 0.48%
20. would: 0.62%	20. societies: 0.48%

Figure 4.7 (March 12, 1947: Truman Doctrine) and figure 4.8 (November 6, 2003: Remarks on Freedom in Iraq and Middle East)

When looking at it through the keywords, we get a stronger sense of how they deal with it. In our collocations we could see that Truman wants to work abroad to help and protect America by working abroad and supporting abroad, but we can see from the keywords how he plans on being supportive, this is by sending financial and economic assistance to help both Greece and Turkey. Meanwhile, this issue is tackled completely differently by Bush, he does not focus on how he will support these countries but still focus on the principles of the state and what they hope to spread into the middle east.

Next, we will look at JFK's speech below in Figure 4.9. This speech as we said from our collocates was intended to inform the nations of the threat to the United States, we can still see this in the speech with keywords like "soviet", "weapons", "nuclear" and "threat". These words along with together words inform the people, but one change from his collocates is that we can see him using "cuba", at first, we did not see it maybe to bring the blame away from them, but they hold as much blame in this speech as the soviets with a 1.50% extraction rate. Still this speech is to indicate the movements of the weapons and the threats they pose to the United States, but from his keywords extracted we see the addition of blame and trying to hold these nations accountable for their actions in the hemisphere.

In Barack Obama's speech below in Figure 4.10, we see a continuation of his policy of normalizing relations with Cuba, with keywords like "future", but it does not appear as prominent in the keywords instead we see words that discusses the people and the nations with words like "people", "cuban", "cuba", "states", and "america" with a few other examples in there, unlike his collocates we see him trying to relate the people of Cuba and the people from the United States, to outline how close they are and how close they can be in the future with their relations. It shows his single approach to their future, unlike the collocates no single words appear that discuss the past of both nations which show that, when it came to keywords, he did not want to make them about the past and instead focus on what is ahead for the two nations and their "future".

1. soviet: 1.93%	1. people: 2.91%
2. cuba: 1.59%	2. cuban: 1.96%
3. weapons: 1.17%	3. cuba: 1.81%
4. missiles: 1.09%	4. states: 1.26%
5. nation: 1.09%	5. united: 1.21%
6. hemisphere: 1.01%	6. believe: 0.85%
7. world: 1.01%	7. cubans: 0.80%
8. nuclear: 0.92%	8. american: 0.75%
9. threat: 0.92%	9. many: 0.70%
10. offensive: 0.84%	10. world: 0.65%
11. nations: 0.84%	11. know: 0.65%
12. government: 0.75%	12. time: 0.60%
13. military: 0.75%	13. president: 0.55%
14. action: 0.75%	14. every: 0.55%
15. united: 0.75%	15. future: 0.55%
16. peace: 0.75%	16. america: 0.50%
17. free: 0.67%	17. years: 0.50%
18. people: 0.59%	18. thats: 0.50%
19. one: 0.59%	19. today: 0.45%
20. buildup: 0.50%	20. ive: 0.45%

Figure 4.9 (October 22, 1962: Address on the Buildup of Arms in Cuba) and figure 4.10 (March 22, 2016: Remarks to the People of Cuba)

Overall, from looking at their keywords we can see how different both JFK's and Obama's approach is to this global issue. For JFK, like his collocates we can see that he was informing the people of the threats from abroad, but from his keywords we see that he puts the blames on both the "soviets" and "cubans" and with Obama we do see "Cuba" mentioned but in a better light, for Obama it is about the "future". About their relations in the futures and from his keywords we understand that he is clearly trying to relate to the Cubans and relate the

Americans with the Cubans to increase reliability and ensure that both nations can become friends.

Next, we will look at Ronald Reagan's speech below in Figure 4.11. This speech focuses on the people still and presents it still as that "national" "crusade" which appear in the collocates and now still in the keywords. It is clear his goal is to relate this issue to everyone and try to get the nation to help each one of the affected people. Keywords like "us" and "together" continue to inspire and inhibition this nature of togetherness In the people, he also uses words like "freedom" and "life" to highlight the things that people on drugs do to have as they would be addicted and forced to love precariously. He clearly wants to inspire people to fight their addictions but also to work together to do so.

1. us: 1.78%	1. thank: 1.82%
2. drugs: 1.71%	2. drug: 1.82%
3. drug: 1.42%	3. going: 1.72%
4. children: 0.85%	4. people: 1.59%
5. new: 0.85%	5. dont: 1.35%
6. one: 0.78%	6. get: 1.31%
7. america: 0.71%	7. want: 1.17%
8. nancy: 0.64%	8. well: 0.84%
9. abuse: 0.64%	9. also: 0.79%
10. see: 0.64%	10. right: 0.70%
11. time: 0.57%	11. us: 0.70%
12. people: 0.57%	12. see: 0.70%
13. together: 0.57%	13. country: 0.65%
14. national: 0.57%	14. drugs: 0.65%
15. help: 0.57%	15. many: 0.65%
16. special: 0.50%	16. come: 0.65%
17. much: 0.50%	17. really: 0.56%
18. freedom: 0.50%	18. much: 0.56%
19. life: 0.50%	19. tough: 0.56%
20. crusade: 0.50%	20. opioid: 0.51%

Figure 4.11 (September 14, 1986: Speech to the Nation on the Campaign Against Drug Abuse) and figure 4.12 (March 19, 2018: Remarks on Combating the Opioid Crisis)

Donald Trump takes a different approach to his speech on the opioid crisis above in figure 4.12. In his collocates, one could see that the speech did little to inspire but focused more on aspects to blame like gangs and the border to the south. In his keywords we see no keywords on fighting it together, which compares so differently to Reagan's speech. There is as much reference to the crisis in the key words such as "drug", "drugs" and "opioid" but in Donald Trump's speech they feel more judgmental when compared to words like "tough" which just indicate the wrongness of the actions. Words like "us" only appear 0.70% of the speech compared to Reagan's 1.78% use of the word. Donald Trump clearly does not see this as that

“national crusade” but more as a crisis, and what I mean is that he sees the wrong, and that it can only be overcome by “tough” action.

This internal issue for Reagan is about community and pulling together to do what’s right and to work together to overcome these issues apart of a “national crusade”. He does not wish to penalize or lay blame on people, but emphasis what things they can miss out on like “freedom” and “life”. Meanwhile Trump deals with this issue in a much different way, he does not have any keywords that could put huge emphasis on a community effort instead, he paints no strong picture of how he wants to deal with this crisis, he use of words in relation to drugs is much the same as Reagan’s but then besides that his does not have any keywords that show any great deal of understanding of the crisis or even a solution.

From looking at the keywords, once again we see an evolution in how Cold war presidents and modern-day presidents tackle their shared global and internal issues by looking at the keywords. Firstly, we see that Truman is looking to protect abroad by supporting his allies economically and financially, meanwhile Bush is about protecting America by action abroad and then justifying the actions in his speeches, this is a massive change from Truman, as it is not a unifying approach to global relations. JFK in his speech seems to inform about also put the blame on the Cubans and the Soviets as they appear at the top of his keywords, but Obama then in relation to Cuba does not blame, instead focuses on relating to the cubans to help the process of normalizing, which shows that over time, the conflict, and crisis has been exchanged for peace and resolution. Now with Reagan, it comes to be just a unifying a approach to the drug crisis, he wants to emphasise that the people must work together to resolve this crisis, but instead Trump goes for a different approach, his approach seems to be empty of understanding or a solution, in his keywords is only reference to the drugs and nothing around the people other than “us” which appears very little, in the end, he does not rally the nation nor give a proper solution to the situation. As stated in my collocation’s conclusion, the cold war seems to be more unifying and that point stands here, modern day speeches have changed but not in a more meaningful way, other than Obama, both Trump and Bush’s own personality shines in these speeches they do not unify, but instead justify and blame, much like JFK but even then, he looks to inform and have solutions. Both Truman and Reagan then want to unify his people and rally their allies against communist and rally their people against drugs.

4.1.3 Sentiment Analysis

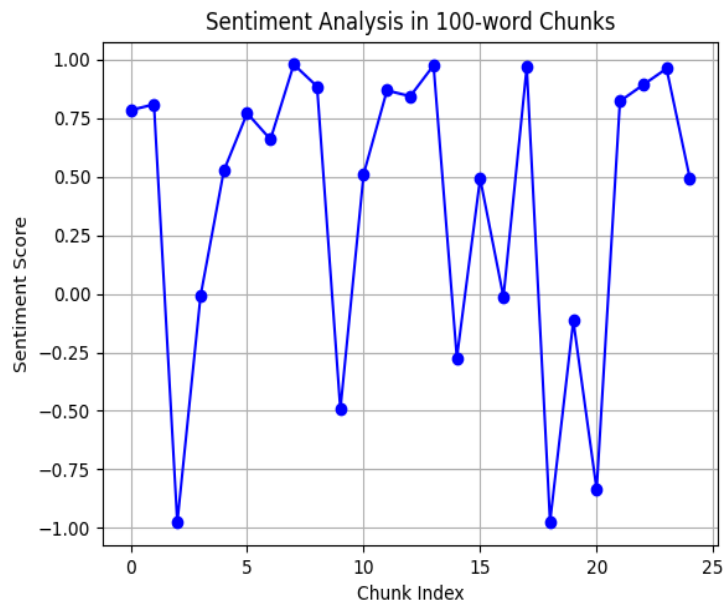


Figure 4.13 (Harry Truman March 12, 1947: Truman Doctrine)

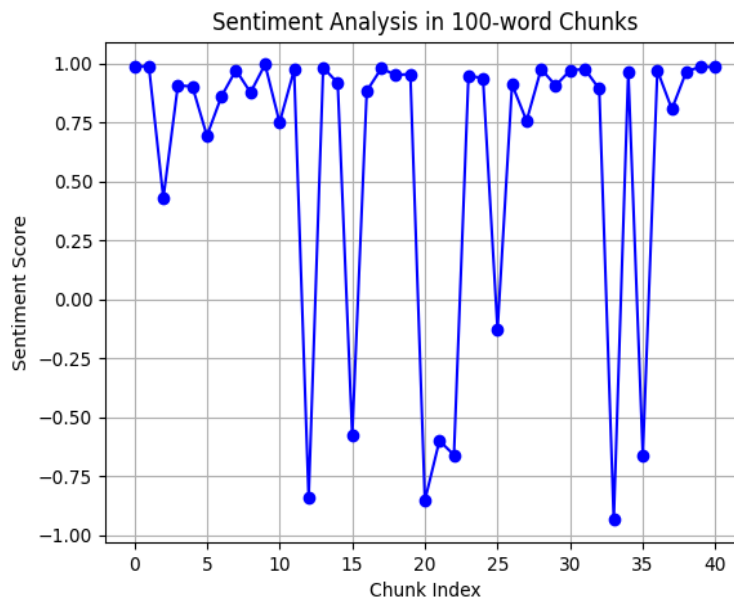


Figure 4.14 (George W. Bush November 6, 2003: Remarks on Freedom in Iraq and Middle East)

Firstly, we will look at the sentiment from both Truman and Bush in their speeches in Figures 4.13 and 4.14 above. First glance we can see that Bush's speech is the far longer speech with over 40, 100-word chunks compared to Trumans 25, 100-word chunks. Both speeches get to the same negative sentiment with both dropping to -1.00 sentiment, but overall, with the

lengths considered, Truman's does appear to be more negative than Bush's as it drops below the neutral line 6 times throughout with Bush dropping below 8 times. We can also see a small pattern develop in the way both speeches are done. Both speeches indeed stay positive at the start, but we then see them take a sporadic individual take, with them dropping, but we see also see at different points in the speeches we see them both drop for 3, 100-word chunks and almost exactly follow the same pattern, we see this in Trumans speech in the end and Bush's speech in the middle. Other than this, both speeches seem to follow similar patterns, but Trumans does appear more negative than Bush's.

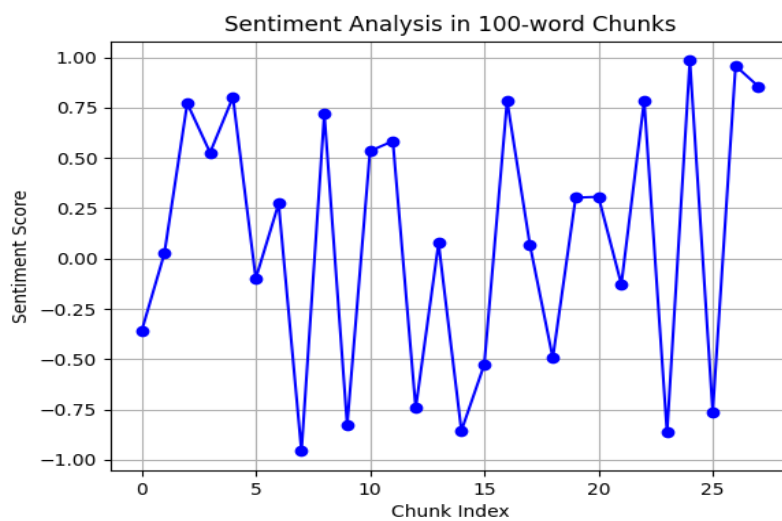


Figure 4.15 (John F. Kennedy October 22, 1962: Address on the Buildup of Arms in Cuba)

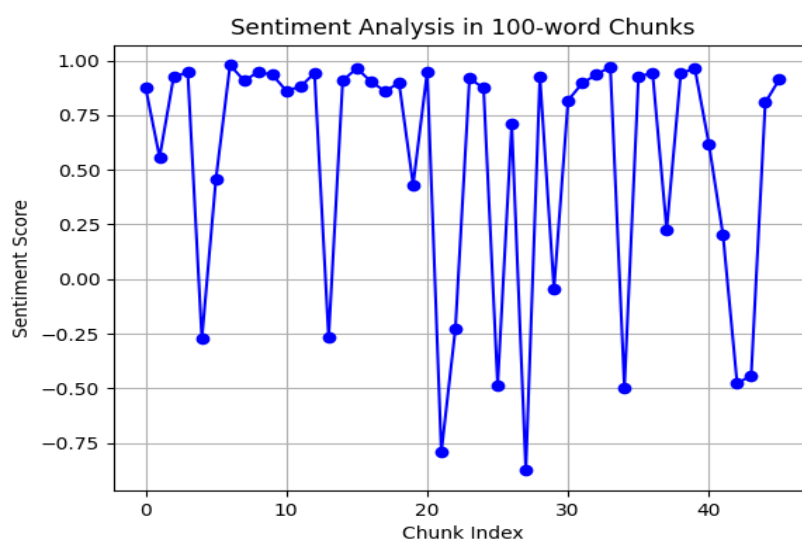


Figure 4.16 (Barack Obama March 22, 2016: Remarks to the People of Cuba)

Now let's look at both JFK and Obama's speeches above in Figures 4.15 and 4.16. Once again, we can clearly see that the modern-day speech by Obama, is far longer than JFK's speech, with 46, 100-word chunks and in JFK's only 28. Only JFK's speech drops to a -1.00 negativity sentiment score, meanwhile Obama's get to only -0.75, apart from this we see once again the cold war speech, is far more negative than the modern say speech. Firstly JFK, drops an equal amount of 9 times, but due to the length of Obama's speech, it ends up as the more negative speech. We also see that Obama's speech is closer to 1.00 positive score will JFK's only reaches that peak 2 times. In terms of patterns or trends, they do not resemble each other in many ways, the only being the positive end to their speeches.

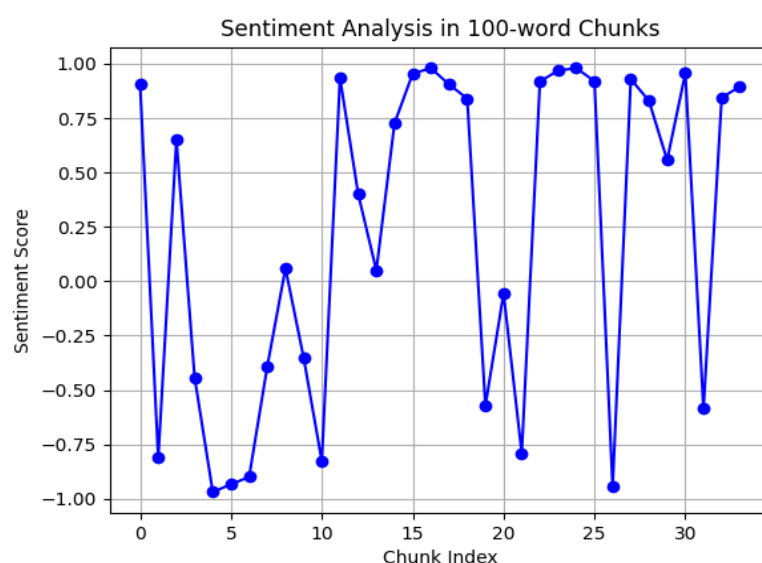


Figure 4.17 (Ronald Reagan September 14, 1986: Speech to the Nation on the Campaign Against Drug Abuse)

3550

Finally, we shall look at Reagan's speech which is above in Figure 4.17 and Trumps in Figure 4.18 below. Donald Trump's speech ends up with 50, 500 words speeches meanwhile Reagan's is only 35, this shows once again that his speech ends up being far longer. In terms of sentiment both speeches drop to -1.00 in their negativity score, but we see that Donald Trump ends up with the more negative speech overall with more then 18, 100-word chunks going bellow the neutral line, alongside this we see that most drop and stay at -1.00, meanwhile Reagan's speech is more positive and does not drop -1.00 as much as Trumps. They do not follow much of the same patterns but one that is striking is that once they drop in sentiment around the start, they both stay at a similar neutral tone for an extended period,

showing that most of the negative sentiment parts of the speech are tackled at the start. Also we see them both ending at a positive sentiment score, and that Donald Trump's stays more positive as the speech comes to its end.

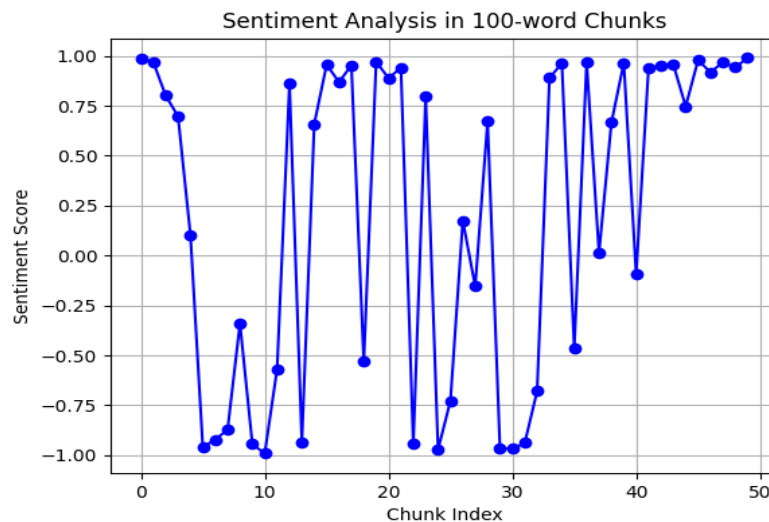


Figure 4.18 (Donald Trump March 19, 2018: Remarks on Combating the Opioid Crisis)

In the end we can see some interesting facts from all the presidents. Firstly, we see that modern day speeches that tackle the same global and internal issues appear to be much longer with all 3 modern day presidential speeches being longer, then to counter this almost the cold war speeches appear to be more negative, with Truman and JFK having a lower negativity score, the only outlier being Donald Trump who is much more negative than Reagan on his speech about drugs. None of the speeches follow an overall pattern or design, showing that for these speeches it depends on the president for the layout of the speech.

4.2 Conclusion

When looking at the collocates we get this very simple style from cold war presidents to modern day presidents when dealing with their shared global and internal issues, mainly that we can notice that the cold war unity is swapped from modern day divergence when approaching the issues. There is more action and individualism in modern day speeches, but cold war is action through community of nation and people. Then with the keywords extracted from these speeches we see this same point shine through, Truman and Reagan emphasis community, with Truman focusing on giving aid to nations in need and Reagan about joining a national crusade against a crisis. Meanwhile Trump and Bush, only blame external and internal factors for the opioid crisis and justifying actions in the middle east.

These do not unite, but instead sow division. Only JFK and Obama take a different approach as JFK looks to inform and to hold those accountable and Obama looks to allow for the normalization of both Cuba and America's relation. When it comes to the sentiment analysis carried out on these speeches, we see that modern-day presidential speeches addressing similar global and internal issues tend to be longer, contrasting with cold war speeches that appear more negative, particularly with Truman and JFK showing lower negativity scores.

Chapter 5

5.1 Analysis of final research question

The aim of this chapter is to analyse and answer the final research question: "What key phrases emerge from Democrat and Republican presidential speeches when analysed using Craig Zeta analysis?" Firstly, an analysis of the words preferred by the Democrats will be conducted. Following this, an analysis will be carried out of the words preferred by the Republicans. Finally, I will compare and conclude on the results to determine whether Craig Zeta can work in tandem with political science.

America operates under a two-party system, where both the Democrats and Republicans compete for power in important elections, such as for Governor, Congress, Senate, and most significantly, for the President of the United States. This position is not only the most powerful in the US but also arguably in the world. The power struggle between the parties has led to significant divisions in the nation, and they rarely see eye to eye. It will be interesting to see how both parties handle their speeches.

5.1.1 Words Preferred (Democratic)

The list generated by the Craig Zeta analysis is quite telling of key phrases from the Democratic party presidents in the dataset. There is a significant mention of Cuba, which may be skewed due to two comparison speeches being based around Cuba. Other than this, the rest of the data gives great insights.

The first word in the list is 'united' this is not only a nod to the nation's namesake, as the United States you would find this to be used a lot. But interestingly below 'united' is not 'states' that takes the 3rd position but instead 'war' is placed below. Firstly, the word 'united' was not only used in terms of the United States but also in reference to the United Nations, which is a key term in Truman's Inauguration speech for example "This should be a cooperative enterprise in which all nations work together through the United Nations and its

specialized agencies whenever practicable”. Mentions of the United Nations, try to bring all the countries together, and shows a more United world approach to the problems faced by the world. Most notably, the Soviet Union, in which 2 of the democratic presidents faced. This is why ‘war’ is so close to the top. As stated before, Truman became president at the end of the second world war, which saw the downfall of Nazi Germany, and the rise of the Soviet Union as the major threat to freedom of the free western nations of the west (U.S. Presidents during the Cold War, 2022). With the Soviet Union as an important player in world politics, Truman was keen on limiting the influence and the spread of communism. Much of the same of John F. Kennedy, who was the closest to all-out war with the Soviet Union, with the Cuban missile crisis. This closely ties to the use of the word ‘nuclear’ that appears on the list of words preferred, as the Cuban missile crisis nearly ended in all out nuclear war, as JFK says in his Cuban missile address “We will not prematurely or unnecessarily risk the costs of worldwide nuclear war in which even the fruits of victory would be ashes in our mouth--but neither will we shrink from that risk at any time it must be faced”. Both the words ‘nuclear’ and ‘war’ appear here, both Truman and JFK explain for the high usage of the term’s ‘war’, ‘nuclear’, ‘soviets’ and ‘armaments’ as these times against the Soviets, were times where war was a fear of many citizens of the west.

Obama, was the only President to not use the term ‘war’ but did mention the wars in Iraq and Afghanistan by saying “We will begin to responsibly leave Iraq to its people and forge a hard-earned peace in Afghanistan.” This point perfectly feeds into the next key phrase, that we see from the democratic presidents. This word is ‘peace’, with so many uses of words like ‘war’ and ‘nuclear’ with democratic presidents, there needed to be what I would consider to be a counterbalance, in the speeches. We see this word used in JFK’s Cuban missile address, “We have no wish to war with the Soviet Union-for we are a peaceful people who desire to live in peace with all other peoples” and furthermore in Trumans doctrine speech “The free peoples of the world look to us for support in maintaining their freedoms. If we falter in our leadership, we may endanger the peace of the world and we shall surely endanger the welfare of this Nation.”, This shows how important this word is, and how as Truman sees it, how important the United States is for word peace when he says above, that if they falter, the world might lose peace. This comes, with much responsibility for the leaders of the United States, they believe themselves to be the country that keeps the world stable and at peace, but at the same time, ‘war’ is a word used so many times in these democratic speeches which is

very contradictory and this contradiction is a heavy one as other words such as ‘offensive’, ‘defence’ and ‘conflict’ add to this list of words preferred.

These 3 democratic presidents, seem to have had a bigger focus on the global tensions at the times leaving little space for the economy. Nonetheless we see some use of 3 key words that are connected to the economy which are ‘economic’, ‘assistance’ and ‘prosperity’. When it comes to the word assistance, this would mostly be becoming from Trumans, doctrine speech, which focuses on sending aid to Greece, during its civil war. With comments like this “Greece is in desperate need of financial and economic assistance to enable it to resume purchases of food, clothing, fuel and seeds”. During both Trumans and JFK’s time in office, there was economic stability and growth, but during Obama’s prosperity had declined, and their country was during a great recession. Obama states that “The success of our economy has always depended not just on the size of our Gross Domestic Product, but on the reach of our prosperity; on our ability to extend opportunity to every willing heart not out of charity, but because it is the surest route to our common good”. This prosperity is the reach of which the Americans can trade and at this point that reach encompasses most of the nations on this earth. Surprisingly these troubles of economic pressure on the United States does not shine through as much at the key phrases around ‘war’, which seems to not take presidency over economy during the presidency of both John F. Kennedy and Harry S. Truman, but for Barack Obama, far outweighed the wars abroad.

Overall, I can take one main key phrases from the speeches of the democratic party presidents that is examined in this report and that is that Democratic or at least these 3 United States Presidents, have been either taking office or are in office during times of great global tension and foreign wars. This is greatly true to both Truman, who was the first president during the cold war and would see the great deterioration of US and Soviet relations after the second world war and JFK who was president during the time where the world was closest to nuclear destruction. Meanwhile Obama took office, with 2 unpopular wars abroad but mainly a time when the economy was hurt. This in mind, we see that these presidents have seen conflict and great global tension, but Barack Obama was the only of the 3 presidents to see great economic hardship. This is why we can see very little mention of the economy. Even with these mentions of economy, we understand that many mention of ‘economic’ and ‘assistance’ appear in Trumans doctrine speech with comments like this “The United States must supply this assistance. We have already extended to Greece certain types of relief and economic aid

but these are inadequate”. Overall, we can see that these presidents key phrases, focus on the tensions with the Soviets and the economy and economic assistance it can provide.

5.1.2 Words Preferred (Republican)

Much like the democrat dataset there is a huge mention of 2 words, this is in relation to Ronald Reagans speech on the drug crisis and Donald Trump’s on the opioid crisis in America. This is why ‘drug’ and ‘drugs’ top the craig zeta analysis. After understanding why these words top the list, the words below give us great number of key phrases to analyse and understand.

When looking at the first few words that have been generated, an interesting image of the Republican party is created. The first words on the list include ‘women’, ‘children’ and ‘citizen’. These words stand out, as it interesting that words relating to 2 main groups and then referencing citizens, shows that the Republican party is trying to relate to the American people. It is used in speeches, referring to ‘men’ and ‘women’ but it is interesting as the word ‘men’ does not appear on the start of the list. Interestingly we see a huge mention of the key phrase ‘women’ when George W. Bush address on freedom in the Middle East and Iraq. For example, he says “In many Middle Eastern countries, poverty is deep, and it is spreading, women lack rights and are denied schooling”, this is the reason we see it placed so high, but this alone is not just the reason it is so high on the list. Seemingly in all republican speeches the term ‘men and women’ is used interchangeably. Like it is a key phrase for all speeches, to relate to all the people in the United States.

Another key phrase I identified was ‘children’. We see this word used heavily by Trump when discussing the opioid problem and then again in Ronald Reagan’s Inauguration speech. For example, Reagan says “For decades we have piled deficit upon deficit, mortgaging our future and our children's future for the temporary convenience of the present”. It is a powerful term, used by all three Republican presidents in their inauguration speeches and in all their comparative speeches. This shows that a key phrase of the Republican party is about appealing to the idea of the American family. Their duty to their ‘children’ to not only drive change in American policy as Reagan declares in his quote above, but also to stop drug use as declared here by Trump in his opioid speech “Hundreds of years never been like this. And we will raise a drug-free generation of American children.” We see the appeal to the idea of the American family once again as further in the list the key phrase ‘families’, we see this word used by Ronald Reagan in his inauguration when discussing the American hero

“They're individuals and families whose taxes support the government and whose voluntary gifts support church, charity, culture, art, and education. Their patriotism is quiet, but deep”.

Furthermore, we can see another key phrase appear in the generated list ‘middle’. This comes in with what we have noticed this far that the republicans are relating to the people, by mentioning ‘middle’ this is in reference to what is the biggest class in america the ‘middle’ class. In the United States more than 50% of all americans are middle class as of now, but during some of these presidential terms it was 61% in 1971 and 59% in 1981 (Frankenfield, 2023). Even with this margin sinking, it makes sense why this is an important class to relate to, for example in Donald Trump’s inauguration speech he says “The wealth of our middle class has been ripped from their homes and then redistribute further across the entire world. But that is the past.” Another word used in tandem with ‘middle’ is ‘working’. America also has a huge working class. During Reagan’s time, the working class was mainly a democratic support base which is why this class is not mentioned in his inauguration, but this has much changed since then (Ruffini, 2023). For George W. Bush we see this class mentioned as they started to drift to the republican party as Bush says in his Inauguration speech “And we will reduce taxes, to recover the momentum of our economy and reward the effort and enterprise of working Americans”. Now this shift is complete, as Ruffini says the realignment of working-class voters helped Trump get into office in 2016 and nearly allowed him to win in 2020 and will be important in 2024 for his campaign (Ruffini, 2023). We can expect more key phrases to arise now that this is a growing voter base for the republican party.

More key phrases lead to the idea of republicans being against “big government”. This term refers to how much influence the federal government has on the day-to-day lives of American citizens (Goddard, n.d). We see this mentioned with words such as ‘establishment’ and ‘federal’. For some presidents this is a key issue, such as when Ronald Reagan says in his inauguration speech ‘It is my intention to curb the size and influence of the Federal establishment and to demand recognition of the distinction between the powers granted to the Federal Government and those reserved to the States or to the people’. This shows that Reagan had a problem, with the federals establishments authority and was keen in returning this power back to the states and the people to manage their affairs. These terms are also a away to give out about big government but also to challenge the government last in power, which we can see Doanld Trump say in his inauguration “The establishment protected itself, but not the citizens of our country. Their victories have not been your victories; their

triumphs have not been your triumphs; and while they celebrated in our nation's Capital, there was little to celebrate for struggling families across our land.”.

Overall, we can see two main focal points come from republican key phrases. The first being the need to relate to the American people and secondly this notion of big government. We see this need to relate with key phrases such as the ‘middle’ and ‘working’ class and with mention of ‘women’ ‘children’ and even ‘families’ to relate to the pride and passion behind the average American family. Even more we see mentions both ‘federal’ and ‘administration’ as two main key phrases to discuss and attack big government that the democratic party applies on the people and states. Surprisingly there is little key phrases to do with the economy of the United States, the only mention is ‘jobs. This is important for all presidents, but it is interesting that little mention of the economy appears, and it only appears in the sense of ‘jobs’, we see this key phrase used by Ronald Reagan in his inauguration speech “There are entrepreneurs with faith in themselves and faith in an idea who create new jobs, new wealth and opportunity.”. As I said this comes as a surprise, as you would expect more mentions of the economy, in what is a peaceful time for the American nation.

5.1.3 Comparison and contrasting

When looking at how both Republican and Democratic presidential speeches compare and contrast. Firstly, they compare, by not having key phrases that reference the economy in a high volume in their speeches. Secondly, they contrast in 2 main ways, the Democratic speeches chosen for this analyse have key phrases that discuss ‘conflict’ and ‘war’ in a such huge volume and these key phrases do not appear in Republican speeches. Meanwhile, in the Republican speeches we see key phrases that are meant to relate to the population and the classes of the american population, meanwhile there is little to no key phrases like this in the democratic speeches.

As discussed in the words preferred list which is the democrats we see very few mentions of the economy and in the words avoided list which is the republicans we once again see that the most important key phrases are not around economy or at least a major part of either presidential speech. In a country with the biggest economy in the world it is surprising that neither the Democrats nor the Republicans make it a major part of their speeches. The 3 Democrat presidents had different economic pressures. When Truman, became president it was in a time when the economy was healed from the depression and stronger than ever (Hamby, n.d) for JFK, the economy had a recession in 1958 and unemployment rate was at around 6% (John F. Kennedy on th economy and taxes, n.d), even with this economic trouble,

the United States were still strong and for JFK trouble with the USSR was more pressing. Meanwhile under Obama, as states the recent recession was fresh in the mind of most Americans and important point of politics in the states. By looking at these 3 presidents, one can understand that for Truman and JFK more importance was to put to the global struggle of the Soviet Union and Barack Obama bring more key phrases of economy to the list including the key phrase of ‘prosperity’ which for him is about how far America’s influence can reach as his says here in his inauguration speech “The success of our economy has always depended not just on the size of our Gross Domestic Product, but on the reach of our prosperity”, meanwhile Truman bring two key phrases with his mentions of ‘economic’ and ‘assistance’ in his speeches but this is not to do with the US economy but relates to what Obama says that the success of the US economy is about the “reach of their prosperity”.

Meanwhile the economy for the Republican presidents was much different and yet much the same. Ronald Reagan came into office at a time of period of economic stagflation (Kenton, 2023) and at a time when relations with the Soviet Union had become friendlier, George W. Bush came into office with a small 2001 recession which had unemployment at 5.7 in that same year (Amadeo, 2021) but had his mind on global tensions with a war on terror on the horizon and Donald Trump came into office with the United States economy in a healthy condition, with unemployment at 4.7% with an annual GDP growth of 2.5% (Kurtz & Yellin, 2020) with all this in mind the only key phrase referencing the economy is ‘jobs’. When comparing why both parties do not discuss economy, I think it comes down to be able to compare all 6 presidents. Ronald Reagan and Barack Obama both come into office when there is a time of economic trouble, George W. Bush and John F. Kennedy mirror each other in that they have small economic troubles but pressure from the outside world is more pressing for them to discuss about. Truman and Trump inherit an economy built by the former president and able to be built upon it is in a strong place in the world physically. Simply put, all the presidents have far more pressing issues then the economy except for Obama and Reagan.

This is where we can see differences, we see key phrases all relate to two main differences, this is that the democrats have more key phrases on ‘war’ and that the republicans have more on key phrases that can relate to the American people. As stated in the Democrat presidential speeches there is many key phrases that appear around ‘war’ and ‘conflict’ meanwhile for the Republicans there are no key phrase words around conflict or war in the Republican list. One can simply look at the dataset and understand that with the Democrat presidents we have

Truman and JFK dealing with tensions of conflict with the Soviet Union and Obama dealing with a war in Afghanistan and Iraq, meanwhile the Republican presidents like Reagan has tension with the Soviet Union but not to the extent of JFK and Truman, George Bush has no conflicts abroad that take main stage for the Americans, as even the Soviet Union is dismantled and the war on terror had not started fully in the speeches chosen. Donald Trump has only to deal with wrapping up the war in Afghanistan. Simply looking at it, might just hold the answer in the dataset but I think its deeper than just that and I believe this is a choice made by the party. This belief comes from the list of the key phrases used by the Republican Party they have far more use of words that relate to the American people with words 'middle', working to relate to the 2 main classes in America and 'women', 'children' and 'families' to relate to the American family and the pride around it. I believe this choice is deliberate, for both parties it is about re-election and staying in power in all institutions of the United States. I believe these parties have different approaches, the Republican parties focus more on the people, to relate and make themselves a party of the people, a party for middle and working class citizens and the preserver of the image of the American family meanwhile the Democrats can be seen as the defenders of democracy not just a party for the people in the US but also the world and the party that will defend the interest of America abroad. This difference is from what I can see, intentional. Every US president has the power to either relate to the people or be seen as a defender of democracy and of the western world, but from these speeches and Craig Zeta analysis I believe it is the party lines that decides this.

5.2 Conclusion

Overall, the idea of using Craig Zeta analysis when it comes to political science is a sure thing. I found it to be powerful for me to be able to get the context of the speeches and the different approaches of both the Republican party and Democrat party. I feel this can be used in further research into politics and is something that can grow our knowledge of the leaders that represent their respective countries and even local speeches but one of the things I would say go hands in hands with this, is understanding context of the speeches and context of the situation of the country that the speechmaker is in, which means more research not only into the speeches but the issues and triumphs of the country. If I had a larger scope for the project a larger study done on more speeches, but also more focus speeches would have yielded more telling results. What I mean by more focused is a single set of speeches such as inaugural speeches and by larger, I mean from all the United States Presidents from both parties. Understanding this, I know that this was a great place to start with Craig Zeta, as it did allow

me to understand the importance of these tools and the great insights we can get from looking at these speeches, in the end it is something that can be used in future research not only in its most used area of literature research but now also in the realm of political science.

Chapter 6

Overall Conclusion

Now that all the analysis is done, let's conclude this report. Firstly, NLP has been a powerful tool in this analysis, it has given this report the very tools to complete all analysis including collocation, keyword extraction and sentiment analysis, this report shows how useful these are when analysing any piece of literature of any kind, including political speeches. Much of the same can be said about Craig Zeta analysis, this is a tool that is underutilized for political science, it has allowed this report to delve deep into the speeches and to understand how both parties seem to handle their speeches and what key points appear in them all, it is a tool that I hope will become more well known as time goes by, to ensure that we understand the words of our politicians and parties. We see from cold war inaugural speeches a great change from cold war to modern day, not only are modern day inaugural speeches shorter but also more positive, but besides this we see them all have their own issues to deal with, but cold war speeches seem to be more inspiring and unifying, comparing to modern day speeches, which almost lose that key point, and the grand scale seems to become more narrow, as the world takes backstage and the rise of more focus on internal policy arose. The keywords show a similar change, as we see the use of more words referencing the "american" people take centre stage, again showing the change to domestic policy.

When it comes to cold war presidents and modern-day presidents tackling global and shared issues, we see more changes, the unity which is shown in the cold war speeches, is changed to divergence once again in the modern-day speeches, the keywords show the same change, that cold war focuses on Community meanwhile Modern day mostly seems to focus on blame and justification abroad. We see a change in terms of sentiment, that modern day speeches, tend to be longer, which shows that while inaugural speeches for the cold war are on average longer, the speeches on specific issues tend to be shorter compared to modern day, but cold war speeches, do still tend to be more negative when facing these issues.

When it comes to looking at republican and democratic speeches, we see some interesting findings from the Craig Zeta analysis. We see a look of Democratic emphasis on "war" and "conflict" meanwhile the Republicans look at relating with words like "middle class",

“working class”, “women”, “children” and “families”. This seems to be a deliberate choice, as one can see the democrats focusing on portraying themselves as the defenders of democracy at home and abroad, meanwhile the republicans want to be seen as heroes of the people, mainly the working, middle class, and American families.

Bibliography

- All, M. (2023, March). *Datacamp*. Retrieved from NLTK Sentiment Analysis Tutorial for Beginners: <https://www.datacamp.com/tutorial/text-analytics-beginners-nltk>
- Amadeo, K. (2021, December 25). *George W. Bush Administration Policies and Impacts*. Retrieved from The balance: <https://www.thebalancemoney.com/bush-administration-economic-policies-3305556>
- Barber, R. (2021). *Early Modern Digital Review (Vol. 4 No. 4 (2021): Special Issue, Computational Approaches to Examining Early Modern Texts)*. Retrieved from Barber, Ros. "Craig's Zeta test.": <https://jps.library.utoronto.ca/index.php/emdr/article/view/38839/29714>
- Barney, N. (2023, March). *sentiment analysis (opinion mining)*. Retrieved from Tech Target : <https://www.techtarget.com/searchbusinessanalytics/definition/opinion-mining-sentiment-mining#:~:text=Sentiment%20analysis%2C%20also%20referred%20to,a%20product%2C%20service%20or%20idea>.
- Beliga, S. (n.d). *Keyword extraction*:. Retrieved from Langnet: https://www.inf.uniri.hr/images/datoteke/web/2014/beliga_kvalifikacijski_final.pdf
- Bigram vs Trigram: Decoding Common Word Mix-Ups*. (n.d). Retrieved from the content authority: <https://thecontentauthority.com/blog/bigram-vs-trigram>
- Bird , S., Klien , E., & Loper, E. (2009). *Natural Language Processing with Python*. Retrieved from <https://tjzhifei.github.io/>: <https://tjzhifei.github.io/resources/NLTK.pdf>
- Boghe, K. (2020, March 12). *Kaggle*. Retrieved from Presidential Speeches - Scraped from Miller Center: <https://www.kaggle.com/datasets/kboghe/presidentialspeeches>
- Brockopp, D. Y. (1983, Jul 7). *The American Journal of Nursing, Vol. 83, No. 7 (Jul., 1983)*. Retrieved from JSTOR: https://www.jstor.org/stable/3463336?casa_token=DXxbXXZdp44AAAAA%3AsbitolokldliowKhan5PSqN0Lv19r7F832D_3Gu5-Ifu1sRp0hfmwwcYKOWVGnNxlMBS8x9tB0C1_2CYv4RRG2Y31wXkefvjx6HEy9Adv0oMFUzargl
- Caetano, J. A., Lima, H. S., Santos, M. F., & Marques-Neto, H. T. (2018, Sep 3). *Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election - Journal of Internet Services and Applications volume 9, Article number: 18 (2018)*. Retrieved from Springer Open: <https://jisajournal.springeropen.com/articles/10.1186/s13174-018-0089-0>
- Chen, X., Yan, Y., & Hu, J. (2019, April 6). *A Corpus-Based Study of Hillary Clinton's and Donald Trump's Linguistic Styles*. Retrieved from Research Gate: https://www.researchgate.net/publication/332248438_A_Corpus-Based_Study_of_Hillary_Clinton's_and_Donald_Trump's_Linguistic_Styles
- Duignan, B. (2024, Mar 12). *Britannica Money*. Retrieved from financial crisis of 2007–08: <https://www.britannica.com/money/financial-crisis-of-2007-2008>
- Eder, M., & Rybicki, J. (n.d). *Stylo R Package*. Retrieved from European Association For Digital Humanities: <https://eadh.org/projects/stylo-r->

- Kenton, W. (2023, January 10). *Reaganomics: Definition, Policies, and Impact*. Retrieved from Investopedia: <https://www.investopedia.com/terms/r/reaganomics.asp>
- Keyword*. (n.d.). Retrieved from Cambridge Dictionary: <https://dictionary.cambridge.org/dictionary/english/keyword>
- Kurtz, A., & Yellin, T. (2020, Oct 28). *economy performed under Trump versus prior presidents*. Retrieved from CNN Business: <https://edition.cnn.com/interactive/2020/10/business/us-economy-trump-vs-other-presidents/#:~:text=At%20the%20start%20of%20Donald,highs%2C%20and%20so%20were%20stocks>.
- Matplotlib: Visualization with Python*. (n.d.). Retrieved from matplotlib: <https://matplotlib.org/>
- McKeown, K. R., & Radev, D. R. (1997, Sep 3). *Collocations*. Retrieved from CiteSeerx: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=479d9c0308f824eb27fd586f1071e4322598ba45>
- Nasukawa, T., & Yi, J. (2003, Oct). *Sentiment analysis: capturing favorability using natural language processing*. Retrieved from ACM Digital Library: https://dl.acm.org/doi/abs/10.1145/945645.945658?casa_token=d0RHO5kvPj0AAAAA:7yCMXfIXksX7zyZmjnnb3kln2TzXNrXlaihBYK5T9ETPnklq16ggj7t7FPdpT1OIsPt_9-awVXcKQ
- NLTK*. (2023, Jan 02). Retrieved from Natural Language ToolKit: <https://www.nltk.org/>
- Rijk, L. d., Derks, L. A., & Grimley, B. (2019). *The evidence for NLP*. Retrieved from International Association for: https://www.ia-nlp.org/pdfdocs/SciRes_TheEvidenceForNLP_LisaDeRijk_etal_2019.pdf
- Ruffini, P. (2023, April 11). *The Emerging Working-Class Republican Majority*. Retrieved from Politico: <https://www.politico.com/news/magazine/2023/11/04/new-republican-party-working-class-coalition-00122822>
- Sentiment Analysis: Comprehensive Beginners Guide*. (n.d.). Retrieved from thematic: <https://getthematic.com/sentiment-analysis/#:~:text=Sentiment%20analysis%20is%20used%20to,or%20%E2%80%9Cemotion%20artificial%20intelligence%E2%80%9D>.
- Seretan, V., Nerima, L., & Wehrli, E. (n.d.). *Extraction of Multi-Word Collocations Using Syntactic Bigram*. Retrieved from <https://ruig-gian.org/ressources/wehrli-art2-Nerima-Seretan-Wehrli.pdf>
- Singh, P., Sawhney, S. R., & Kahlon, S. K. (2017, oct 4). *Forecasting the 2016 US Presidential Elections Using Sentiment Analysis*. Retrieved from Springerlink: https://link.springer.com/chapter/10.1007/978-3-319-68557-1_36
- U.S. Presidents during the Cold War*. (2022, Jul 13). Retrieved from Richard Nixon Foundation: <https://www.nixonfoundation.org/2022/07/u-s-presidents-cold-war/#:~:text=Harry%20Truman&text=At%20the%20beginning%20of%20the,to%20Europe's%20post%20war%20recovery>.
- War on Drugs*. (2024, March 5). Retrieved from Britannica: <https://www.britannica.com/topic/war-on-drugs>

What Is an N-Gram? (n.d). Retrieved from MathWorks:

<https://www.mathworks.com/discovery/ngram.html#:~:text=An%20n%2Dgram%20is%20a,t ext%20classification%2C%20and%20text%20generation>.

What is R? (n.d). Retrieved from r-project: <https://www.r-project.org/about.html>

Yadav, K. (2021, Aug 28). *5 Ways to Generate Colocations/N-Grams*. Retrieved from Medium:

<https://towardsdatascience.com/generating-colocations-n-grams-6dd8bea31d13>