

Presidential Inaugural Address Analysis

Luke Chesley, Caleb Miller, Harrison Pauwels

Click here to access our dashboard - <https://final-633-b74277a80624.herokuapp.com/>

Abstract

How has the language of United States presidents changed over time? By analyzing each inaugural address in the history of the country, we can uncover interesting insights into shifting tendencies that reflect not only America as a country but also the geopolitical trends within. This paper describes the findings from an exploration of every inaugural address delivered by United States presidents. Many different facets of language were investigated to uncover a narrative that spans more than 230 years of history using a wide range of tools.

Introduction

Purpose

The purpose of this work is to explore and visualize the results of analysis on presidential inaugural addresses. These speeches were chosen as the subject of analysis because of the circumstances each speech is delivered under. A newly elected president of the United States addresses the people he represents and speaks about the state of the nation, the state of the world and his goals for the upcoming term. Each speech, except for the first one, has taken place on March 4th or March 5th in the year following the presidential election at the Capitol building. The consistent environment, magnitude and audience of these speeches make them a cohesive data set

to analyze. These speeches reflect America and the American people and by discovering trends in these speeches we can also gain insight into American history.

Dataset

This data was gathered from The American Presidency Project, <https://www.presidency.ucsb.edu/>, which is a project created and maintained by UC Santa Barbara. They house many different types of presidential documents such as eulogies, executive orders, state of the union addresses and inaugural addresses. A web scraping program was used to obtain the inaugural addresses.

Methods

I. Obtaining and Cleaning Data

This stage of the research was conducted entirely in Python using the requests and BeautifulSoup modules. Each address is housed on a unique page on the American Presidency Project. It was necessary to create a loop for each page url to scrape the contents with the requests module. After scraping, the contents were then parsed with BeautifulSoup. The raw text of the speech was extracted along with the date of the address and name of the president speaking. Some non-text characters, including new line and tab characters, were captured during the scraping process so steps were taken to remove them leaving just the raw text.

II. Analyzing Data

Python string manipulation functions were used to do simple analysis of the text. The features obtained with this method include the number of words in each speech, the characters per word and the number of unique words.

NLTK was used for more advanced analysis such as determining the average number of words in each sentence as well as determining what part of speech each word is. NLTK was also used to find and remove 'stop words'. Stop words are words that appear very often in language and do not add much meaning on their own. Excluding these words can help uncover and clarify subtleties in the data that were not otherwise visible. The default NLTK stopwords list was used throughout this research. During analysis, the data was housed and manipulated in Pandas dataframes.

The Flesch-Kincaid score that appears throughout the research is a widely used method of measuring the readability of English language passages. It was developed by the United States Navy in 1975 to assess the difficulty of technical manuals and is now used by many states as a regulatory tool. For example, Pennsylvania requires automobile insurance policies to maintain a score below a certain score to keep them accessible to all of their customers. The formula used to derive the score takes into account the total number of words in a passage, the total number of sentences and the total number of syllables. The exact formula is as follows:

$$206.835 - 1.015(\text{total words} / \text{total sentences}) - 84.6(\text{total syllables} / \text{total}$$

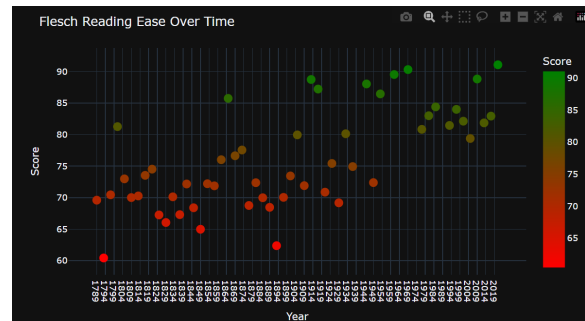
words). The scores can be interpreted with the help of the chart in appendix A.

III. Graph and Dashboard Creation

We were interested in creating a collection of coherent graphs that were easily viewable and interactive to make visual decoding of the overarching narrative as clear as possible. The graphs used throughout this report were created with the Python module Plotly. These graphs were made interactive with the Python module Dash. By combining the functions of these libraries we can deliver the viewer the ability to filter the graphs by presidents and years. By default, the dashboard includes all data from all inaugural addresses but by adding the filtering functionality we create the opportunity for fast, ad hoc analysis by each unique viewer on any subset of the data that is of particular interest to them.

Data Exploration

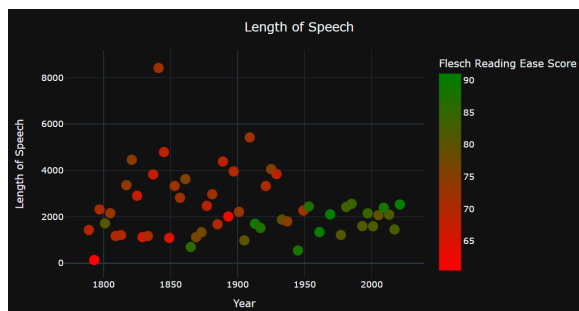
Flesch-Kincaid



The graph tracing the Flesch-Kincaid Reading Ease scores of presidential inaugural addresses over the centuries reveals a significant evolution in the complexity of language used. The initial scores, especially in the 18th and early 19th centuries, hover around the lower 60s, which

indicate texts that are considered fairly difficult to understand. This complexity suggests that the language of the time was formal and perhaps less accessible to the average listener. A pivotal change occurs starting around 1949, where the scores begin an upward trajectory, signaling a shift toward simpler, more accessible language. For example, one can note a pronounced leap in the scores from around 65 in the early 1800s to scores that sometimes exceed 80 in speeches given after 1949. By 2019, the trend appears to consolidate, with scores consistently in the higher range, reflecting a clear intention to make these pivotal addresses more understandable to the general populace. This strategic shift in presidential rhetoric aligns with a modern imperative to communicate effectively with a diverse and broad audience.

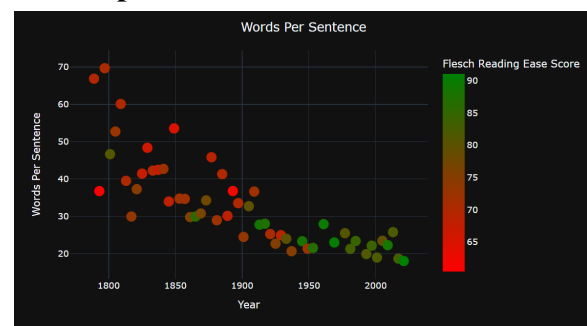
Length of Speeches



Presidential inaugural speeches typically average between 2,000 and 3,000 words in length. Recently, there has been a trend towards these addresses becoming shorter and more straight forward. Before 1950, significantly longer speeches, like William Henry Harrison's 8,425-word address, were not unusual. Nowadays, it appears that candidates have identified an optimal length for these speeches and are consistently sticking to it.

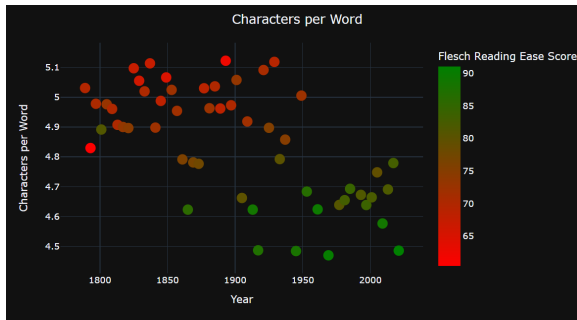
Stop words, which are frequently used but don't contribute significantly to the meaning of a sentence (like 'a', 'and', 'the', etc.), are an interesting aspect of language. When these stop words are removed from presidential inaugural speeches, the overall length of the speeches is reduced by half. Despite this, the trend of these speeches maintaining an optimal length remains consistent.

Words per Sentence



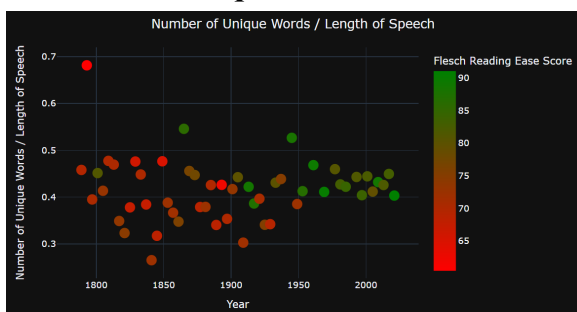
There has been a notable decrease in the number of words per sentence in presidential inaugural speeches. For perspective, George Washington's first speech was below average in terms of length at 1,430 words. Yet, the average length of his sentences, at 66.87 words, was among the highest in history and was reflective of the speaking style of his era. In comparison, Joe Biden's inaugural speech, which was a bit lengthier than usual for recent times at 2,532 words, had the lowest average words per sentence of any U.S. president, at just 18 words per sentence.

Characters per Word



The data reveals a steady decline in the average number of characters per word in presidential speeches over the centuries. In the early 1800s, speeches featured words with an average of 5.1 characters, pointing to a more complex lexicon. Over time, this complexity has given way to simplification, with the modern era averaging a succinct 4.6 characters per word. This subtle yet telling trend underscores a linguistic shift towards simplicity, possibly reflecting an intention to enhance the comprehensibility of presidential addresses.

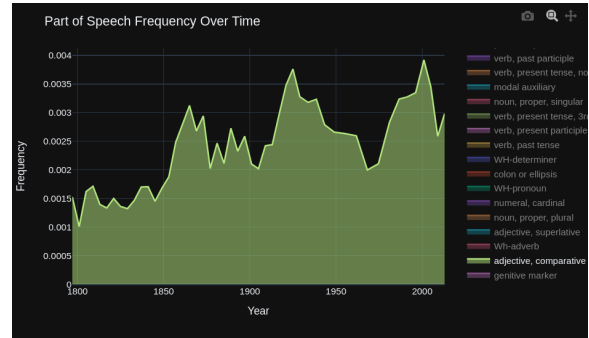
Number of Unique Words



George Washington's second inaugural address stands out for its exceptionally high percentage of unique words, at 68.15%, which is unmatched by any other address. This remarkable feat can be attributed to the very below average length of his speech, which contained only 135 words (making it an outlier). When examining all other presidential addresses, however, the

percentage of unique words remains relatively constant over time, hovering around 40%.

Part of Speech Frequency Over Time



The graph analyzing the parts of speech used in the addresses was found to be noisy in its initial state. A five speech moving average was added to smooth out the graph and make the trends clearer. It can be difficult to discern information when all parts of speech are viewed together so it is necessary to isolate individual parts of speech to extract information. Isolating comparative adjectives a clear trend emerges. The frequency of comparative adjectives has more than tripled since the year 1800 from 0.01% of words to approximately 0.03% of all words spoken. While this is a fairly subtle increase when considering all words, it becomes more significant when considering what the meanings of these words are. Comparative adjectives are words that compare two entities like larger, faster, higher and better. In the context of inaugural addresses, recent presidents have become more accustomed to comparing their own goals, accomplishments and vision for the country than their predecessors have. This observation pairs with the overall trend of decreasing complexity in these speeches.

also offer practical value for political figures in crafting messages that resonate.

Conclusion

The data visualizations of the linguistic trends in presidential inaugural speeches illuminate a strategic shift towards more direct, accessible communication with the electorate. Design choices in these visualizations, such as color coding and chronological plotting, effectively highlight the historical trajectory of speech complexity, underscoring the trend of candidates speaking more plainly over time. Employing elements of the Gestalt principles such as proximity and similarity, these design elements facilitate a clearer understanding of how political figures are engaging with a populace that now, more than ever, values the personal touch over the formality of political speechmaking. These visualizations serve as a compelling narrative tool, charting the evolution of political communication in a way that is as informative as it is visually engaging.

References

Kincaid, J. Peter; Fishburne, Robert P. Jr.; Rogers, Richard L.; and Chissom, Brad S., "Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel" (1975). *Institute for Simulation and Training*. 56. <https://stars.library.ucf.edu/istlibrary/56>

Appendix

Score	<u>School level (US)</u>	Notes
100.00–90.00	5th grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0–80.0	6th grade	Easy to read. Conversational English for consumers.
80.0–70.0	7th grade	Fairly easy to read.
70.0–60.0	8th & 9th grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0–50.0	10th to 12th grade	Fairly difficult to read.
50.0–30.0	College	Difficult to read.
30.0–10.0	College graduate	Very difficult to read. Best understood by university graduates.
10.0–0.0	Professional	Extremely difficult to read. Best understood by university graduates.