Main Narrative Notebook: U.S. Presidential Inaugural Addresses

Navein   Clara   Calvin   Halasya

Thursday 18th December, 2025

Curvenote

1 Overview and Purpose

This notebook serves as the primary narrative synthesis for our project analyzing U.S. Presidential Inaugural Addresses. The purpose of this notebook is to integrate, summarize, and interpret the results obtained across multiple analysis notebooks, with a particular emphasis on exploratory data analysis, descriptive text statistics, sentiment-based measures, and thematic patterns. Rather than introducing new computational methods, this notebook focuses on explaining what was found, why it matters, and how the results relate to broader historical and political contexts.

This project uses the full corpus of U.S. presidential inaugural addresses, spanning from George Washington's first inauguration to the most recent presidency available at the time of data collection. Inaugural addresses represent a unique and historically consistent genre of political speech: they are delivered at a fixed institutional moment, serve a ceremonial and symbolic role, and are intended to communicate national values, priorities, and unity. Because of this consistency, they provide an ideal dataset for examining how political rhetoric evolves over long periods of time.

1.1 Research Questions

The analyses conducted in this project are guided by the following questions:

- What are the most common words used in inaugural addresses, and how do these reflect the priorities of presidential rhetoric?

- How do word counts and character lengths of inaugural addresses vary across time?

- Are there systematic differences between first-term and second-term inaugural addresses in terms of word count and length?

- How has the tone of inaugural addresses changed over time?

- Has there been an increase in fearmongering or polarizing language over time?

- Which president delivers the most polarizing inaugural address according to sentiment-based metrics?

- What are the most common themes present in inaugural addresses?

- How have these themes changed across historical periods?

- Does party affiliation appear to be associated with more polarizing rhetoric?

These questions are addressed using exploratory text analysis methods rather than formal hypothesis testing, with the goal of identifying broad patterns and trends rather than making causal claims.

2 Data Description and Construction

The dataset used in this project was created programmatically using a reproducible web-scraping pipeline implemented in `make_data.py`. This script uses HTML parsing tools to scrape inaugural address transcripts from the American Presidency Project at the University of California, Santa Barbara. The scraped HTML content is cleaned, parsed, and transformed into a structured CSV file suitable for analysis.

Each observation in the dataset corresponds to a single inaugural address and includes the following core variables:

- `president_name`: Name of the president delivering the address

- `president_number`: Numerical order of the presidency

- `date`: Date of the inaugural address

- `text`: Full transcript of the inaugural address

From the raw text, additional variables were derived in subsequent analysis notebooks, including word counts, character counts, sentence-level statistics, sentiment scores, and theme-related metrics. These derived features enable both descriptive analysis and higher-level interpretation of rhetorical patterns.

3 Assumptions Underlying the Analysis

Because this project relies on scraped historical text data, several assumptions are required to justify the analysis:

1. **Textual Accuracy**
   We assume that the inaugural address transcripts hosted by the American Presidency Project accurately reflect the speeches as delivered. While minor transcription differences may exist, these are assumed not to systematically bias the analysis.

2. **Consistency of Formatting**
   The speeches span more than two centuries, during which transcription norms and stylistic conventions have evolved. We assume that variations in punctuation, capitalization, and paragraph structure do not materially affect higher-level text statistics such as word counts, character lengths, or sentiment scores.

3. **Comparability Across Time**
   We assume that it is meaningful to compare textual features across historical periods, while acknowledging that norms of political speechwriting and public communication have changed substantially over time.

4. **Exploratory, Not Causal**
   All results presented in this project are descriptive and exploratory. We do not claim causal relationships between time, party affiliation, political context, and speech characteristics.

These assumptions are standard for exploratory text analysis and allow us to focus on identifying long-term patterns in presidential rhetoric.

---

4 Results and Interpretations

5 Part 1: Exploratory Data Analysis, Word Usage, and Speech Length

The first part of the analysis focuses on exploratory data analysis and descriptive text statistics. The goal of this stage is to understand the basic structure of the dataset and establish baseline patterns before making interpretive comparisons.

5.1 Common Words in Inaugural Addresses

Analysis of word frequency reveals that inaugural addresses consistently emphasize collective identity, national unity, and governance. Frequently occurring words include references to the nation, the people, democracy, freedom, and shared values. This finding aligns with the ceremonial purpose of inaugural addresses, which are intended to unify the country and articulate a vision for the presidency.

While these words appear across nearly all speeches, their relative prominence varies over time, suggesting that although the core rhetorical themes remain stable, the emphasis placed on different concepts evolves.

5.2 Word Counts and Character Length Over Time

One of the most striking results from the exploratory analysis is the wide variation in speech length across U.S. history. Early inaugural addresses are substantially longer, with some containing several thousand words. In contrast, modern addresses are considerably shorter and fall within a narrower range of word and character counts.

This pattern suggests a long-term shift in presidential communication style. Early presidents often treated the inaugural address as a formal written document aimed at elite or institutional audiences. Over time, the address has become more concise and accessible, likely reflecting changes in media technology, audience attention, and the increasing importance of broadcast communication.

5.3 First-Term vs. Second-Term Inaugural Addresses

Comparisons between first-term and second-term inaugural addresses reveal systematic differences in length. Second-term addresses tend to be shorter on average than first-term addresses, both in word count and character length.

This difference may reflect the distinct purposes of the two types of speeches. First-term inaugurals often introduce a president's vision and priorities, whereas second-term inaugurals may emphasize continuity, reflection, or legacy. The observed length differences suggest that rhetorical goals shift once a president has already served a term in office.

6 Part 2: Tone, Sentiment, and Polarization Over Time

Part 2 of the analysis moves beyond descriptive statistics to examine the emotional tone and rhetorical intensity of inaugural addresses.

6.1 Changes in Tone Over Time

Sentiment analysis indicates that inaugural addresses are generally neutral to positive in tone, consistent with their ceremonial and unifying function. However, meaningful variation exists across historical periods.

Some addresses exhibit more optimistic or aspirational language, while others adopt a more cautious or urgent tone, often corresponding to periods of national crisis or transition. This suggests that presidents adapt the emotional framing of their rhetoric to historical context while maintaining the overarching goal of unity.

6.2 Fearmongering Language Over Time

Figure 3 shows a clear long run increase in fearmongering language across U.S. presidential inaugural addresses. While early inaugurals exhibit relatively low and variable rates of fear related terms, the smoothed trend line rises noticeably beginning in the late 19th and early 20th centuries and remains elevated through the modern era. Although individual speeches fluctuate, the overall pattern suggests that fear oriented framing becomes more prevalent in later decades.

Importantly, this increase does not imply that inaugural addresses are dominated by fear based rhetoric. Rather, it indicates that presidents increasingly acknowledge threats, crises, or national challenges as part of their framing, often in conjunction with calls for unity or collective responsibility. The trend is gradual and cumulative, consistent with changing historical contexts and communication norms rather than abrupt rhetorical shifts.

6.3 Polarizing Language Over Time

In contrast, Figure 2 provides little evidence of a sustained increase in polarizing language over time. Rates of explicitly polarizing terms remain low across the entire time span, with the smoothed trend line staying relatively flat and exhibiting only modest short term fluctuations. While a few early speeches show isolated spikes, these do not persist across adjacent presidencies or develop into a long term trend.

Overall, the graph suggests that despite changes in tone and emphasis, inaugural addresses have largely maintained their unifying rhetorical structure. Even as fear related language becomes more common, this shift does not correspond to a parallel rise in polarization, reinforcing the idea that fear oriented framing and polarization are distinct rhetorical phenomena in this corpus.

7 Part 3: Themes, Time, and Party Affiliation

Part 3 synthesizes insights from earlier analyses by examining thematic patterns and their relationship to time and party affiliation.

7.1 Common Themes in Inaugural Addresses

Thematic analysis reveals recurring themes such as national unity, democracy, freedom, responsibility, and progress. These themes appear consistently across the dataset, reinforcing the idea that inaugural addresses serve a stable ceremonial function.

7.2 Thematic Change Over Time

While core themes persist, their relative emphasis changes across historical periods. Early speeches emphasize constitutional legitimacy and institutional foundations, whereas modern speeches place greater emphasis on collective identity, social challenges, and calls to action.

This shift reflects the changing role of the presidency and evolving expectations of presidential leadership.

7.3 Party Affiliation and Polarization

Finally, the analysis examines whether party affiliation is associated with more polarizing rhetoric. While some differences in rhetorical framing across parties are observable, these differences are generally smaller than the effects of historical period.

This suggests that time and historical context play a larger role than party affiliation in shaping inaugural rhetoric, reinforcing the unifying and institutional nature of the address.

8 Integrated Interpretation

Taken together, the results of this project suggest that U.S. presidential inaugural addresses have evolved toward shorter, more accessible, and more emotionally framed speeches. While the core purpose of unity and legitimacy remains consistent, rhetorical strategies adapt to changing political, social, and media environments.

The findings illustrate how political language evolves gradually over time, shaped by institutional constraints and historical context rather than abrupt partisan shifts.

---

9 Limitations and Future Work

This analysis has several limitations. The dataset includes only inaugural addresses and does not capture other forms of presidential communication. Textual analysis is limited to surface-level features and does not incorporate deeper semantic modeling or causal inference.

Future work could extend this project by incorporating topic modeling, sentiment dynamics at finer temporal resolutions, comparisons with other presidential speeches, or regression-based analyses incorporating historical covariates.

---

10 Author Contributions

Each team member contributed meaningfully to the completion of this project, including data collection and preprocessing, exploratory analysis and visualization, interpretation of results, and project coordination. All team members participated in discussion, revision, and final approval of the narrative and analyses. Specifically:

- **Calvin**: Completed part3.ipynb, which investigated themes over time and party affiliation. Added .gitignore, LICENSE, Makefile. Initialized environment.yaml, created make_data.py and test_make_data.py.

- **Clara**: Completed data analysis in part2.ipynb, involving the evolution of polarization and fearmongering over time using sentiment analysis and graphs. Also setup detailed README.md file.

- **Halasya**: Completed main.ipynb, synthesized result and interpreted findings. Created BibTex file

- **Navein**: Completed data analysis in part1.ipynb, which involved various graph visualizations and word frequency analysis. Also setup the myst website, binder link, and GitHub webpage hosting functionality.

---

11 Reproducibility and References

This project is fully reproducible using the files and code provided in the repository. The dataset was generated using a documented web scraping pipeline, and all analyses were conducted using reproducible Jupyter notebooks. Bibliographic information for the dataset and any additional readings is managed using a BibTeX file (`references.bib`), which is integrated into the project via MyST to ensure consistent and automatically rendered citations.

The primary data source for this project is the American Presidency Project at the University of California, Santa Barbara, which maintains a comprehensive archive of U.S. presidential documents, including inaugural addresses \cite{american_presidency_project_inaugurals}.