

The Aggressiveness of Two of the Most Influential Leaders of U.S. Politics: Trump vs. Biden

Domenico G. P. Solidoro

2025-05-07

Table of contents

Introduction	1
Speech Dataset	2
Dictionary-Based Aggression Measurement	2
Machine Learning Aggression Rating (OpenAI API)	2
Data Processing	3
Visualizing Rhetorical Aggression Over Time	4
Conclusion	5
Appendix	5
Contingency table	5
Graph code	6

Introduction

In contemporary politics, the primary resource through which political figures engage, mobilize, advocate, and even polarize audiences is with the help of public speech. This report examines the presence and evolution of th rhetorical aggression in the speeches of two major political figures in the United States of America as well as in foreign policy: Donald J. Trump, who won 2024 presidential election against Democratic Kamala Harris, and Joseph R. Biden Jr., former U.S. President from 2021-2025. Specifically, it investigates the question: ***How do***

levels of rhetorical aggression in their public speeches compare over time and events? Using Python language and R, this project analyzes how each speaker's tone shift in key political years, with particular attention to moments such as election cycles and national crises. By comparing their patterns, the report seeks to uncover broader trends in political communication which sometimes are not believed to be experienced.

Speech Dataset

To assess rhetorical aggression in political speech, I gathered a dataset of public speeches statements delivered by Donald Trump and Joe Biden between 2015 and 2024. These texts were taken from publicly available presidential archives, campaign websites, and transcripts publish by organizations such as Rev. The Dataset also includes major addresses such as campaign rallies, State of the Union speeches, press briefing, and debates.

For the analysis per se, I combined two complementary approaches: the **dictionary-based method** and a large language model (**LLM**) evaluation using OpenAI API (ChatGPT).

Dictionary-Based Aggression Measurement

The first method relies on a pre-defined aggression-related words, such as **failure**, **furious**, **stubborn** and many more. Derived from a similar *Linguistic Inquiry and Word Count* (LIWC)-style lexicon, it is a widely used tool in text analysis that assigns words to psychological categories, including the frequency of aggression to approximate a trait. I calculated the **aggression score** for each speech by computing the proportion of words in the text that matched any word in the aggression dictionary with the following formula:

$$AggressionScore = \frac{(NumberOfAggressionWords)}{(TotalWordCount)}$$

This approach offers a transparent and easily interpretable results. However, it is limited by its inability to capture sarcasm, context, or subtle rhetorical strategies.

Machine Learning Aggression Rating (OpenAI API)

To fill the void left by the previous method, I used OpenAI API's ChatGPT (gpt-3.5-turbo) to rate the aggressiveness of each speech on a scale from 0 (not aggressive) to 1 (very aggressive). I prompted the model with shortened version of each speech (max. 200 words) and asked OpenAI to assign an aggressive score, returning only the numerical rating. This is how the AI was instructed:

i Prompt

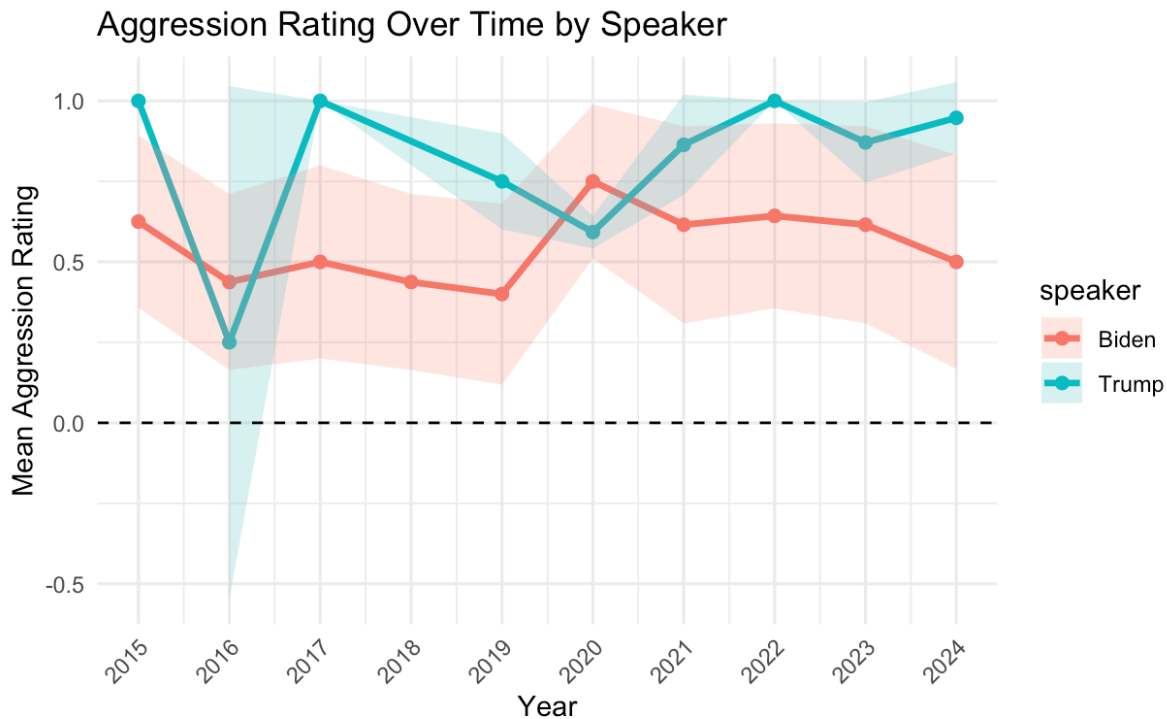
“Rate the following political speech’s rhetorical aggressiveness by assigning 0 (not aggressive) or 1 (very aggressive). Only respond with a single number.”

This classification approach allowed me to capture more nuanced expressions of aggression, such as tone, implication, or figurative language that a simple dictionary method misses.

Data Processing

- All speech texts were cleaned out from any possible deviation;
- I calculated total word counts and aggressive word proportions for each speech;
- I then applied the GPT-based scoring on shortened versions of the speech (only applicable with a paid subscription to generate tokens);
- Finally, I labeled each speech as “*aggressive*” if:
 1. The dictionary-based score was > 0 (i.e., at least one aggressive word);
 2. Or the ChatGPT rating exceeded 0.5.

Visualizing Rhetorical Aggression Over Time



The graph represented above presents the combined aggression scores of speeches delivered by Donald Trump and Joe Biden in a range of time between 2015 and 2024. Each data point reflects a single speech, plotted over time, and classified as either Trump (in blue) and Biden (in red). The y-axis represents the aggression score, taken from the combination of dictionary and OpenAI API evaluations, as described above. A score closer to 1 indicates higher rhetorical aggressiveness.

Several clear patterns emerge from the visualization:

- As expected, Trump frequently registers **high aggression levels**, with noticeable spikes in 2015, 2017, 2022 and 2024 - often tied to **campaign announcements**, early presidency rhetoric, and contentious moments in public discourse. Strategically, his **lowest aggression scores** appear during the key **election years** of 2016 and 2020, suggesting a possible shift to more moderate language during those times.
- Biden, often **perceived as a calm communicator**, shows a slight different result. His **highest aggression score** occurs in **2020**, the year he challenged Trump for the U.S. office. After that, his **tone gradually decreased**, reaching lowest levels by 2024, when both Trump and public opinion framed him as too advanced to run for presidency.

Nonetheless, Biden's aggression is not significantly lower than Trump's overall, and at times, such as in 2020, he even surpassed him.

These findings complicate the simplistic view of Trump as the aggressor in political rhetoric between the two leaders. Instead, they suggested that **both figures strategically escalate aggression in response to political pressures and public expectations.**

Conclusion

This analysis challenges the popular assumption that rhetorical aggression in U.S. politics is a one-sided phenomenon majorly dominated by figures like Donald Trump.

By applying both a dictionary-based method and GPT-generated aggression scores rated from 0 to 1, the study reveals that **Joe Biden also engages in elevated levels of rhetorical aggression**, particularly in politically transitory periods like the 2020 election. While Trump's peaks are frequent and consistent across his public career, Biden's tone demonstrates strategic escalation, showing that aggression is **a tool both leaders use**, not a trait unique to one.

These findings underscore how the style is deeply shaped by **context and timing** - not just personality. As political communication becomes increasingly performative, measuring rhetorical aggression offers a powerful lens to understand how leaders shape narratives, appeal audiences, and contend for power.

Appendix

This is a technical appendix for the operations performed to create this study.

Contingency table

```
print("Trump:")
print(pd.crosstab(trump_df['dict_aggressive'], trump_df['gpt_aggressive']))

print("Biden:")
print(pd.crosstab(biden_df['dict_aggressive'], biden_df['gpt_aggressive']))
```

Trump

Dictionary Aggressive	GPT Not Aggressive	GPT Aggressive
False	71	90
True	99	237

Biden

Dictionary Aggressive	GPT Not Aggressive	GPT Aggressive
False	45	56
True	20	24

The results show how often dictionary-based and GPT-based aggression labels agree or differ for Trump and Biden speeches.

Graph code

```
library(ggplot2)
library(reticulate)

# Pull from Python
aggression_trends2 <- reticulate::py$aggression_trends

# Plot
ggplot(aggression_trends2, aes(x = year, y = mean_aggression, color = speaker)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  geom_ribbon(aes(ymin = ci_lower, ymax = ci_upper, fill = speaker), alpha = 0.2, color = NA) +
  labs(
    title = "Aggression Rating Over Time by Speaker",
    x = "Year",
    y = "Mean Aggression Rating"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = seq(min(aggression_trends2$year), max(aggression_trends2$year))) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  geom_hline(yintercept = 0, linetype = "dashed")
```

With this code, I was able to come up with a “ribbon” graph of the two leaders thanks to the `combined-df` dataset, used to literally combine both `trump_df` and `biden_df` into one.