CBE 1

Amalia, Lauren, and Jenny

2024-10-01

Table of contents

1 Introduction

This project conducts a sentiment analysis to compare the sentiment expressed by Democrats, Republicans, and moderators in presidential debates, focusing on how it has changed over time and how it might differ between parties. We analyze debates from 2012 and 2024 to explore shifts in tone and emotional language across political parties and debate roles. We included moderators in our analysis because the way they frame their questions and how they guide the discussion itself can influence the tone of the debates and how candidates present their messages. This question is interesting as it reveals how changes in debate sentiment reflect shifts in the way politicians communicate and how the different parties may present themselves and then be interpreted by the public through their words, which is especially relevant with the upcoming presidential election and the recent debates.

2 Data

The data for this study consists of the first presidential debate transcripts from Obama vs. Romney (2012) and Harris vs. Trump (2024). These transcripts were obtained online and then divided based on who was speaking, with three groups identified for each debate: Democrat, Republican, and Moderator. For the 2024 debate there were two moderators which were grouped together in a single category. A python script was used to perform pre-processing on the data and divide it properly.

Table 1: Number of Tokens by Speaker

Debate Participant	Total Tokens
Obama	7129
Romney	8300
Harris	8109
Trump	5944
Obama Moderator	1437
Harris Moderators	3106

Table 2: Number of Tokens Spoken by Year

Year	Total Tokens
2012	16866
2024	17159

Tables 1 and 2 summarize the distribution across parties and years. The words spoken by moderators nearly doubled between the two debates, which is partially due to the fact that there were two moderators instead of one. Trump spoke the least out of the four presidential candidates and Romney spoke the most. The 2024 debate was slightly longer, and overall the Republicans talked more than the Democrats.

Table 3: Overall Sentiment Scores of Presidential Debate Speakers

Speaker	Overall_Sentiment
Obama	0.0447752
Romney	0.0494484
Trump	-0.0213891
Harris	0.0498515
Harris Moderator	0.0148697
Obama Moderator	0.0332426

The overall sentiment analysis of the presidential debate transcripts produced the overall sentiment scores for each participant found in Table 3. The results indicate that Obama, Romney, and Harris displayed positive sentiments, while Trump had a slightly negative sentiment. The moderators for both debates showed slightly positive sentiment scores.

3 Methods

First, we compare the emotional valence expressed by Democrats, Republicans, and moderators in the 2012 and 2024 presidential debates over time. We used the sentimentr and text libraries in R for the analysis, cleaning the text data and segmenting it into sentences to calculate sentiment scores.This method was chosen because it effectively captures the shifts in sentiment across different segments of the debates, showing how sentiment can vary not only between candidates but also across time during each debate. It also makes it easier to identify trends and compare emotional responses from the Democrats, Republicans, and moderators.

We also chose to use a Welch Two Sample t-test because it accounts for unequal variances between groups and doesn't require equal sample sizes. This method tests for whether the average emotional valences from the 2012 and 2024 debates are significantly different. Thus, the Welch test can help us draw conclusions about changes in emotional tone across the two years and reveal trends or shifts in how debate participants spoke.

We further categorized our corpuses by part-of-speech - focusing on adjectives since they have the greatest crossover with sentiment. From this, we performed keyness table analysis to identify differences within parties and across years and frequency and dispersion analysis to identify the most common adjectives for each candidate. We included the sentiment denomination of these words according to the "bing", "afinn", "loughran", "nrc" lexicons to determine if a candidate has frequent usage of neutral, positive, or negative language. This method was chosen because it allows us to view the words within the corpuses and draw more specific conclusions across time.

4 Results

4.1 Emotional Valence

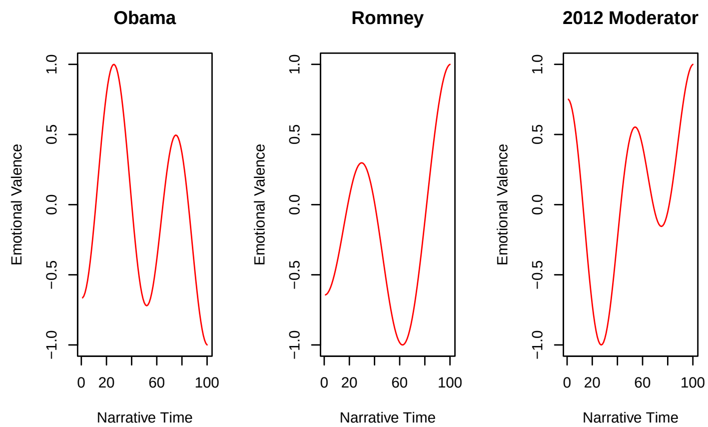

Figure 1: Emotional Valence vs. Time for the 2012 Debate Participants

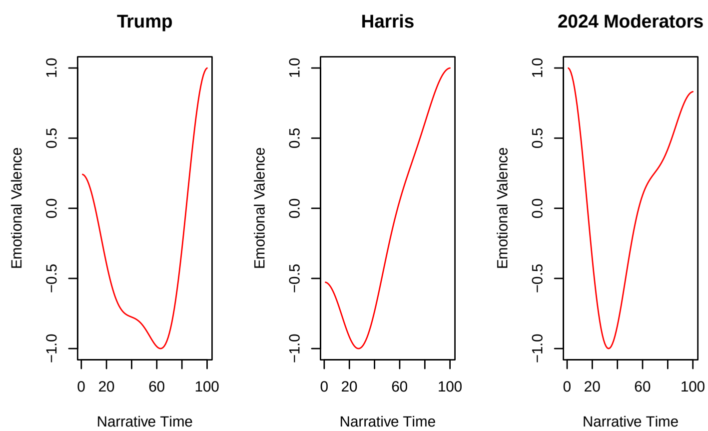

Figure 2: Emotional Valence vs. Time for the 2024 Debate Participants

First, we analyze the emotional valence expressed by Democrats, Republicans, and moderators in the 2012 and 2024 presidential debates over time in Figures 1 and 2. In 2012, Obama, Romney, and the moderator all exhibited fluctuating emotional valences. In contrast, the 2024 debate displayed Trump experiencing a dramatic initial drop in emotional valence, followed by a rise, while Harris reflected similar fluctuations, though less pronounced. The moderators in 2024 showed more dramatic fluctuations than Trump. Notably, all participants in the 2024 debate had similar trends in emotional valence. However, in the 2012 debate, the candidates showed similar patterns that were almost opposite to that of the moderator.

Table 4: Average Emotional Valence by Participant

participant	average_valence
2012 Moderator	0.0584797
2024 Moderator	0.0472646
Harris	-0.1431562
Obama	-0.0117412
Romney	-0.1641919
Trump	-0.3618718

We also found the average emotional valence values for each participant in Table 4. All presidential candidates had a slightly negative average emotional valence value, though Trump seemed to be the most negative. The moderators were slightly more positive in that regard.

4.2 Welch Two Sample t-test

We also conducted a Welch Two Sample t-test to compare the emotional valence of participants in the 2012 and 2024 debates. We found a statistically significant difference (t = 2.2763, p = 0.02319), testing at a level $\alpha = 0.05$, which indicates that the emotional valence of participants in 2024 was significantly more negative (mean = -0.1526) compared to those in 2012 (mean = -0.0392), with a 95% confidence interval for the true difference in means ranging from 0.0156 to 0.2113. Thus, we can reject the null hypothesis that the two debates have equal means (average emotional valences).

4.3 Keyness and Frequency Analysis for Adjectives

We then decided to look at the keyness tables for the Democratic and Republican candidates to determine what words might have been influential in their overall sentiments. We found that adjectives were the most statistically significant, with the most frequent verbs, nouns, and other parts of speech being far more generic and neutral.

Keyness Table

Comparison of Harris to the Obama reference corpus

Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.3_Tar	Per_10.3_Ref	DP_Tar
former	25.37	5.27	0.00000	16	0	2.34	0.00	0.597
middle	7.93	3.60	0.00486	5	0	0.73	0.00	0.903
common	4.76	2.86	0.02917	3	0	0.44	0.00	0.863
only	4.76	2.86	0.02917	3	0	0.44	0.00	0.823
particular	4.76	2.86	0.02917	3	0	0.44	0.00	0.909
tired	4.76	2.86	0.02917	3	0	0.44	0.00	0.939
weak	4.76	2.86	0.02917	3	0	0.44	0.00	0.942
new	3.73	2.60	0.05352	5	1	0.73	0.12	0.848
american	3.39	0.81	0.06576	26	18	3.81	2.18	0.411
big	3.17	2.27	0.07492	2	0	0.29	0.00	0.914

Keyness Table

Comparison of Obama to the Harris reference corpus

Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.3_Tar	Per_10.3_Ref	DP_Tar
sure	43.85	5.15	0.00000	43	1	5.21	0.15	0.700
nuclear	16.87	4.53	0.00004	14	0	1.69	0.00	0.891
safe	7.23	3.31	0.00717	6	0	0.73	0.00	0.948
able	7.05	2.43	0.00791	13	2	1.57	0.29	0.882
long	4.95	2.73	0.02616	8	1	0.97	0.15	0.928
chinese	3.61	2.31	0.05728	3	0	0.36	0.00	0.973
competitive	3.61	2.31	0.05728	3	0	0.36	0.00	0.986
effective	3.61	2.31	0.05728	3	0	0.36	0.00	0.978
glad	3.61	2.31	0.05728	3	0	0.36	0.00	0.966
iranian	3.61	2.31	0.05728	3	0	0.36	0.00	0.983

Evidently, most of the most frequent adjectives for both Obama and Harris are neutral. The Harris data has the negative adjectives *tired* and *weak* as the 6th and 7th most frequent adjectives and the Obama data has the positive adjectives *safe*, *effective*, and *glad* as the 3rd, 8th, and 9th most frequent words. More than that, there appears to be very little overlap between the two sub-corpus in terms of average frequency.

Keyness Table

Comparison of Romney to the Trump reference corpus

Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.3_Tar	Per_10.3_Ref	DP_Tar
sure	36.25	5.71	0.00000	26	0	2.64	0.00	0.835
military	18.64	4.18	0.00002	18	1	1.83	0.10	0.918
nuclear	16.46	2.89	0.00005	22	3	2.24	0.30	0.886
strong	10.93	2.51	0.00095	17	3	1.73	0.30	0.903
clear	6.97	3.33	0.00828	5	0	0.51	0.00	0.958
very	6.97	3.33	0.00828	5	0	0.51	0.00	0.968
right	6.25	3.01	0.01240	8	1	0.81	0.10	0.934
economic	5.58	3.01	0.01820	4	0	0.41	0.00	0.976
necessary	5.58	3.01	0.01820	4	0	0.41	0.00	0.967
peaceful	5.58	3.01	0.01820	4	0	0.41	0.00	0.971

Keyness Table

Comparison of Trump to the Romney reference corpus

Token	LL	LR	PV	AF_Tar	AF_Ref	Per_10.3_Tar	Per_10.3_Ref	DP_Tar
many	15.84	3.99	0.00007	16	1	1.61	0.10	0.719
bad	10.03	2.80	0.00154	14	2	1.41	0.20	0.742
okay	8.27	3.57	0.00403	6	0	0.60	0.00	0.970
weak	8.27	3.57	0.00403	6	0	0.60	0.00	0.881
different	7.30	3.16	0.00690	9	1	0.91	0.10	0.739
horrible	6.89	3.31	0.00866	5	0	0.50	0.00	0.886
big	6.14	2.99	0.01320	8	1	0.81	0.10	0.851
guilty	5.51	2.99	0.01887	4	0	0.40	0.00	0.952
liberal	5.51	2.99	0.01887	4	0	0.40	0.00	0.976
afraid	5.01	2.80	0.02513	7	1	0.71	0.10	0.938

From the Trump and Romney discussion, we saw a similar pattern of a lack of overlap in high frequency words. Interestingly the Romney data's most frequent adjectives have more positive sentiment than negative sentiment; the 4th, 7th, and 10th most frequent adjectives are *strong*, *right*, and *peaceful*. Comparatively, the Trump data's 2nd, 4th, 6th, 8th, and 10th most frequent adjectives are *bad*, *weak*, *horrible*, *guilty*, and *afraid*, which all carry a negative sentiment.

5 Discussion

Most interesting is the difference between the three career politician candidates - Obama, Romney, and Harris - and the non career politician candidate - Trump. Trump appears far more willing to use negative language. Beyond this difference, the sentiment of the candidate's language is more informed by the time period than it is by their party. This could be due to the changing view of politics and politicians by the average person. Even though Harris has an overall more positive sentiment when compared to Trump, the two are more negative than their party predecessors - Romney and Obama.

The validity of our t-test is unclear because of the source data's distribution. Using such a small sample size means that we cannot determine if it is normally distributed. Although it was found that the emotional valence was significantly more negative, the true meaning of this significance cannot be validated from the data in this report.

A future analysis would involve drawing on the transcripts from additional presidential debates. This would allow us to better see trends over time since we would have more than two data points. One example of deeper analysis would be a sentiment and valence comparison across incumbents as a category rather than just parties or years. We could also interpret the tokens in both the context of the original document and the context of the time period - drawing on the fact that many of the tokens in the corpus are informed by the happenings of the world.

6 Acknowledgments

ChatGPT was used to help generate the first draft. This primarily took the form of supplementing our lack of knowledge of different forms of sentiment analysis and certain statistical concepts. We asked the LLM about different forms of sentiment analysis and what libraries should be used to perform said analysis in R. We also used the LLM to develop the code used for initially parsing the text into their separate text files in python. When attempting to fit the Keyness tables on the pdf, I made many queries to the LLM about my problem. In this regard, the LLM was extremely unhelpful as none of its suggestions fixed the problem and they ultimately did not make it to the final product. The LLM was also used to generate a certain amount of regex expressions since I am unfamiliar with the necessary syntax. This ultimately also did not make it to the final report since the analysis in which it was used was deemed unnecessary. Ultimately, the LLM was far more useful in supplementing our knowledge of sentiment analysis than it was in our knowledge/implementation of the R code.

7 Works Cited

1. Brown, D. W. (n.d.). textstat documentation. https://browndw.github.io/textstat_docs/

2. Commission on Presidential Debates. (n.d.). Debate transcripts. Debates.org. https://www.debates.org/voter-education/debate-transcripts/

3. Flynn, L. (2019, June 18). Comparing sentiment analysis dictionaries in R. Medium. https://medium.com/@laurenflynn1211/comparing-sentiment-analysis-dictionaries-in-r-c695fca64326