Campaigning on Twitter: A Comparative Analysis of Clinton and Trump's Linguistic Styles

Aylin Karapanar

14.01.2024

Campaigning on Twitter: A Comparative Analysis of Clinton and Trump's Linguistic Styles

The 2016 US presidential elections were a turning point in US history, an election that would result in the first female president or a president who had no prior experience in public office. While both candidates were unloved and unapproved by the majority of US citizens (Yourish, 2016), the Republican candidate Donald Trump despite losing the popular vote won against the Democratic party candidate, former senator Hillary Clinton by winning 30 states and the decisive electoral college (Beckwith, 2023).

Compared to previous presidential elections, the role of Twitter is undeniable in broadcasting candidates' messages and candidates engaging with voters while providing new opportunities to construct their image and essential information for voters about their campaigns (Buccoliero et al., 2020). This paper is on the tweets covering the months leading up to the election from April until September. This paper aims to answer the question of how the style of tweets for both candidates differed in terms of complexity, lexical richness, and speech patterns, how the style changed over time, and how their tweets represented the candidates' campaigns.

Methodology

The provided dataset included 1399 tweets from Clinton and Trump in the span of the 1st of April 2016 to the 27th of September 2016. The total number of tweets per candidate differed as Clinton had 490 tweets whereas Trump had 909 tweets in the sample. The analysis is done by using Python 3.11.7 (Van Rossum & Drake, 2009) on the Jupiter Notebook application (Kluyver et al., 2016) with the use of Pandas (McKinney et al., 2010), Natural Processing Toolkit (Bird et al., 2009), Scikit-learn (Pedregosa et al., 2011), NumPy (Harris et al., 2020) libraries.

The relevant text pre-processing steps were applied to the dataset. The tweets were tokenized, part of speech tagged, the non-alphanumerical characters were removed, the tweets were lemmatized, and were set to lowercase. After these steps, the following analyses were conducted. The average word count per tweet over months per individual candidate was calculated. The type-token ratio (TTR) was calculated, which is a metric used to express vocabulary richness and diversity of a text quantitatively by using the ratio of types which are unique words occurring in a text by tokens in the text was calculated for each candidate over the months (Wang & Liu, 2018). The most distinctive words by each candidate over the months were extracted from the tweets by using the Term Frequency - Inverse Document Frequency (TF-IDF) method including the removal of English stop words. Part of speech tags n-grams were extracted to evaluate frequent speech patterns and their change over the months. The frequencies of certain pronouns were calculated for each candidate over the months, the frequencies were also turned into ratios respective to the number of tweets in that month for the specified candidate. Lastly, the frequency of first, last, and full names in referring to their opponents was calculated for each candidate.

Results

The average tweet length per candidate fluctuated over the months. The total average number of words in Clinton's tweets was 21.05 words with the lowest value being 20.45 words in July and the highest being 21.79 in September. For Trump, the total average was 20.76 words, the lowest value was 19.34 words in September and the highest was 21.71 in July. The relevant figure can be seen in Appendix A, figure A1.

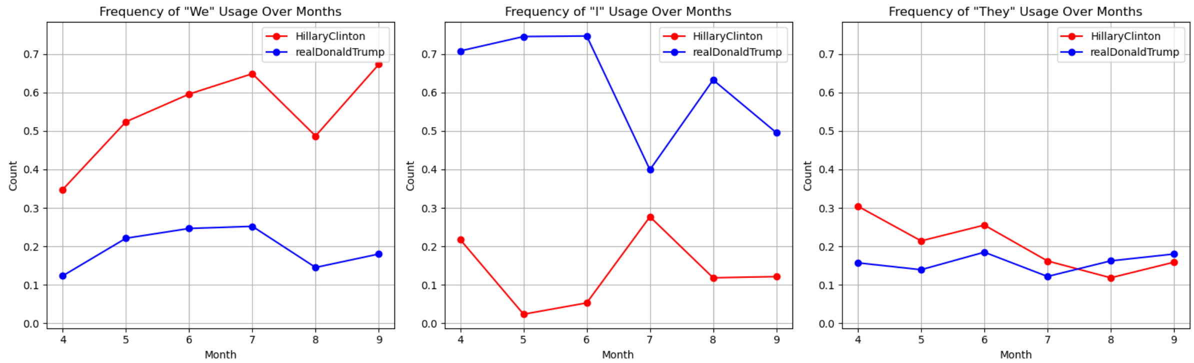
The TTR of each candidate and progression over the months are shown in Appendix A, figure A2. Clinton's tweets' TTR averaged 0.91 with its highest value being 0.92 in April and the lowest was 0.90 in June. Trump's tweets' TTR average was 0.93 with the highest being 0.94 in September and the lowest being 0.92 in June.

TF-IDF method yielded the most distinctive 10 words in candidates' tweets over the months, respective tables for each candidate can be seen in Appendix B. The most frequently used part of speech bigrams for each candidate over the months can be seen in Appendix C.

The frequency of each pronoun group for each candidate and its change over the month can be seen in Figure 1. The highest pronoun to the number of tweets ratio average for Clinton was 'we' pronoun group and for Trump, it was 'I' pronoun group.

Figure 1

Frequency of "I", "We", and "They" in Tweets Over the Months

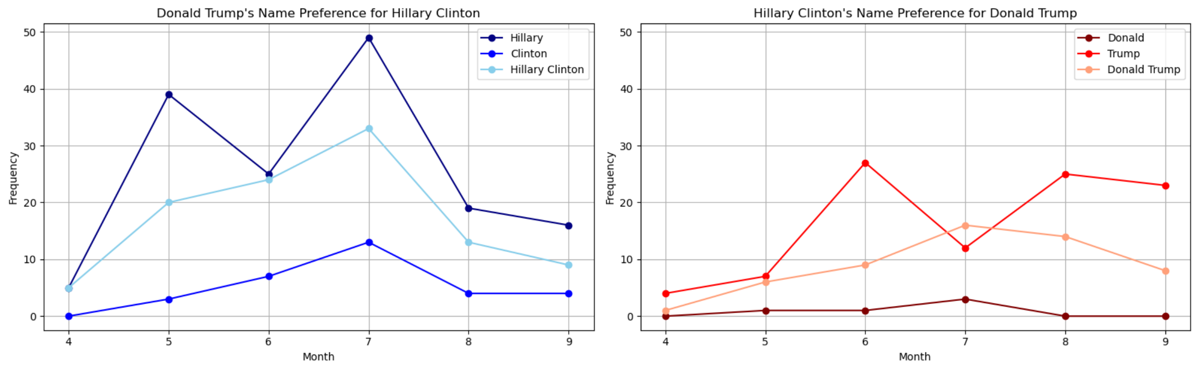


Note. The raw frequencies were divided by the number of tweets by each candidate in that month to get a ratio of frequency relative to the number of tweets.

Lastly, the frequencies of first, last, and full names when referring to their opponents for each candidate over the months are shown in Figure 2. Trump's use of first name for his opponent is more frequent compared to other name options. Whereas Clinton most frequently referred to her opponent by his last name.

Figure 2

First, Last or Full Name Preference for their Opponent of Candidates



Interpretation

The results of the mentioned analyses reflect stylistic and contextual differences between the tweets of the candidates. The average tweet length by words shows a contrasting pattern between the candidates, when the length of Clinton's tweets shows a rising trend, the length of Trump's tweets appears to have a downtrend. The overall averages show that Clinton tended to write longer tweets.

Trump's tweets have a higher average of TTR value compared to Clinton's tweets which can suggest lexical complexity. However prior research on Clinton and Trump's style concluded that Clinton uses more complex language in both her speeches, debates, and tweets (Savoy, 2018; Wang & Liu, 2018). The selected time frame of the tweet could have affected the results, because it only covers five months, and the majority of those months are before the official announcement of both candidates (Roberts & Owen, 2016; Siddiqui et al., 2016). In addition, number of tweets of Trump were higher than Clinton, which could have influenced the TTR values.

Most distinctive words for each candidate reflect their campaign, but this is more apparent in Clinton than in Trump. Most distinctive words for Clinton represent her political agenda such as education, healthcare, women's rights, and equal pay, and include the name of her opponent. Most distinctive words for Trump include names of other politicians, some distinctive phrases include Trump's choice of adjectives describing them and the name of the media outlets. The outcome is in line with the content analysis done by Buccolerio et al. (2020), who investigated the topics and themes in candidates' tweets. They found that almost half of Trump's tweets include attacks on other people almost a quarter of Clinton's tweets were about her political agenda and another quarter of them were attacks on Trump, these patterns can also be seen in the most distinctive words.

The most frequent POS bigrams for Clinton included determiner-noun (DT, NN), noun-noun (NN, NN), and proper noun-proper noun (NNP, NNP) which suggests a diverse and descriptive language. Common POS bigrams for Trump included proper noun-proper noun (NNP, NNP), noun-preposition (NN, IN), and preposition-proper noun (IN, NNP), the prevalence and the pattern of these bigrams suggests a more repetitive language construct. The outcome resembles the candidates' POS patterns in their TV interviews, debates, and speeches which were analysed by Savoy (2018), the results of Savoy's study suggested that Trump used a direct, simple style, and similar expressions that were reused; Clinton used a richer vocabulary covering a wide range of topics while using descriptive rhetoric including nouns, adjectives, prepositions, and determiners.

The use of 'I' was more prevalent in Trump's tweets and the use of 'we' was more frequent in Clinton's tweets, showing differences in communication styles, Clinton tweeted more with the aim of unifying and finding commonalities with the voters hence she frequently used 'we', Trump as discussed previously tweeted to contrast him and his opponents hence he employed 'I' (Buccolerio et al., 2020; Chen et al., 2019). When referring to their opponent, Clinton used his last name and Trump used her first name most often which might suggest a gender bias as Atir and Fergusson (2018) examined that people are more likely to refer to male politicians by surname and female ones by first name. However, it should be noted that this might have been to distinguish Hillary Clinton from Bill Clinton as Hillary Clinton tended to use her first name in her campaign.

Conclusions

In recent elections Twitter became a mass communication tool for politicians, allowing them to broadcast their messages and communicate with their voters. However, the way that Twitter is used differs by each politician and this is evident in the candidates of the 2016 US presidential election. Tweets by Clinton and Trump have been analysed to determine their stylistic and contextual differences, the result suggest that Hillary Clinton's tweets were longer, less lexically complex, more about her political agenda, more diverse in terms of POS patterns, and Clinton employed a more unifying tone with the frequent use of 'we', compared Trump who used shorter tweets and repetitive POS patterns to talk more about his contrast with others while using 'I' more frequently. Their choice in referring to their opponent was different, Trump using first name and Clinton using surname more often. This can suggest a gender bias, but also could be a way to distinguish Hillary Clinton. In conclusion, differences in tweets reflect candidates' linguistic choices, communication styles, and campaign strategies.

References

Ahmadian, S., Azarshahi, S., & Paulhus, D. L. (2017). Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences*, *107*, 49–53. https://doi.org/10.1016/j.paid.2016.11.018

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media Inc.

Beckwith, D. C. (2023, December 12). United States Presidential Election of 2016. Encyclopedia Britannica. https://www.britannica.com/topic/United-States-presidential-election-of-2016

Buccoliero, L., Bellio, E., Crestini, G., & Arkoudas, A. (2020). Twitter and politics: Evidence from the US presidential elections 2016. *Journal of Marketing Communications*, *26*(1), 88–114. https://doi.org/10.1080/13527266.2018.1504228

Chen, X., Yan, Y., & Hu, J. (2019). A Corpus-Based Study of Hillary Clinton's and Donald Trump's Linguistic Styles. International Journal of English Linguistics, 9(3), Article 3. https://doi.org/10.5539/ijel.v9n3p13

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Kluyver, T., Ragan-Kelley, B., Fernando Perez, Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90).

McKinney, W., & others. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51–56).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

Roberts, D., & Owen, P. (2016, July 20). US election 2016: Republican party nominates Donald Trump for president. *The Guardian*. https://www.theguardian.com/us-news/2016/jul/19/donald-trump-nominated-president-republican-convention

Savoy, J. (2018). Trump's and Clinton's Style and Rhetoric during the 2016 Presidential Election. *Journal of Quantitative Linguistics*, *25*(2), 168–189. https://doi.org/10.1080/09296174.2017.1349358

Siddiqui, S., Roberts, D., & Gambino, L. (2016, July 26). Hillary Clinton named presidential nominee as Democrats make history. *The Guardian*. https://www.theguardian.com/us-news/2016/jul/26/hillary-clinton-presidential-nomination-democratic-convention

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Wang, Y., & Liu, H. (2018). Is Trump always rambling like a fourth-grade student? An analysis of stylistic features of Donald Trump's political discourse during the 2016 election. Discourse & Society, 29(3), 299–323.

Yourish, K. (2016, June 3). Clinton and Trump Have Terrible Approval Ratings. Does It Matter? *The New York Times*. https://www.nytimes.com/interactive/2016/06/03/us/elections/trump-and-clinton-favorability.html, https://www.nytimes.com/interactive/2016/06/03/us/elections/trump-and-clinton-favorability.html

Appendix A

Figure A1

Average Tweet Length per Month - Trump vs Clinton.

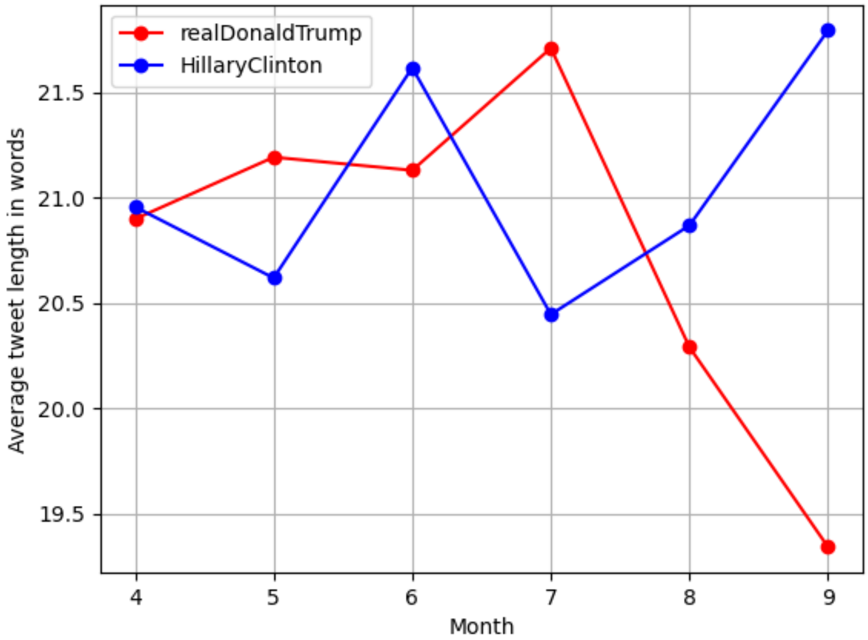
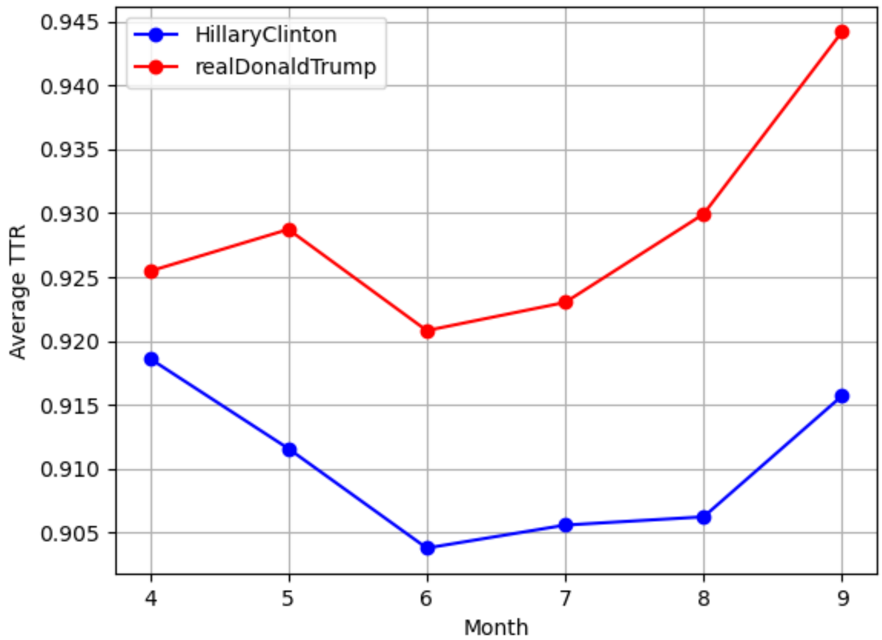


Figure A2

Average Type-Token Ratio per Candidate Over the Months



Appendix B

Table B1

Ten Most Distinctive Words for Hillary Clinton Over the Months

	4	5	6	7	8	9
0	family	trump	trump	trump	trump	trump
1	right	make	make	donald	donald trump	president
2	trump	donald	america	hillary	donald	say
3	woman	family	family	make	family	work
4	people	good	just	america	america	just
5	new	care	let	potus	tax	need
6	pay	work	woman	need	just	people
7	equal	donald trump	need	donald trump	work	america
8	support	pay	donald	let	doesn	country
9	need	right	donald trump	people	people	make

Note. Each column represents the most distinctive words used by the candidate in that given month, 0 being the most distinctive.

Table B2

Ten Most Distinctive Words for Donald Trump Over the Months

	4	5	6	7	8	9
0	great	hillary	hillary	hillary	hillary	hillary
1	cruz	crooked	clinton	crooked	great	great
2	ted	crooked hillary	crooked hillary	crooked hillary	clinton	say
3	kasich	great	crooked	clinton	say	clinton
4	delegate	say	hillary clinton	bernie	crooked	people
5	ted cruz	clinton	bad	hillary clinton	medium	enjoy
6	vote	woman	great	great	people	just
7	lyin ted	hillary clinton	just	people	hillary clinton	thank
8	lyin	thank	make	bad	cnn	foxandfriends
9	hillary	nytimes	people	just	bad	poll

Note. Each column represents the most distinctive words used by the candidate in that given month, 0 being the most distinctive.

Appendix C

Table C1

Three Most Frequently Used POS Bigrams for Hillary Clinton Over the Months

	4	5	6	7	8	9
0	(DT, NN)	(NNP, NNP)	(NN, NN)	(NN, IN)	(NN, IN)	(DT, NN)
1	(NN, NN)	(NN, NN)	(NN, IN)	(NNP, NNP)	(DT, NN)	(TO, VB)
2	(JJ, NN)	(JJ, NN)	(DT, NN)	(JJ, NN)	(JJ, NN)	(NN, IN)

Note. Each column represents the most frequent POS bigrams used by the candidate in that given month, 0 being the most common. NNP means proper singular noun, NN means common singular noun, IN means preposition, DT means determiner.

Table C2

Three Most Frequently Used POS Bigrams for Donald Trump Over the Months

	4	5	6	7	8	9
0	(NNP,NNP)	(NNP, NNP)	(NNP, NNP)	(NNP, NNP)	(NNP, NNP)	(NNP, NNP)
1	(IN, NNP)	(NN, IN)	(NN, IN)	(NN, IN)	(IN, NNP)	(NN, IN)
2	(NN, IN)	(DT, NN)	(JJ, NN)	(IN, NNP)	(NN, IN)	(JJ, NN)

Note. Each column represents the most frequent POS bigrams used by the candidate in that given month, 0 being the most common. NNP means proper singular noun, NN means common singular noun, IN means preposition, JJ means numeral or ordinal adjective.