

Trends in American Political Vernacular

Nicholaus Jackson and Tory Peterschild
Courant Institute of Mathematical Sciences, NYU
New York, USA

Abstract— American politicians must speak to their electorate with the level of linguistic sophistication that the electorate understands. As a result, political speeches are great indicators of the state of the American vocabulary.

Keywords—analytics

I. INTRODUCTION

The premise of this project is that American politicians must speak to their electorate with the level of linguistic sophistication that the electorate understands. As a result, political speeches are great indicators of the state of the American vocabulary. This project then sets out to prove that over the last century, the vernacular used by the average American — representative of the electorate — has been drastically simplified. This will be achieved by doing unique word counts on American political speeches.

II. MOTIVATION

As the 2016 presidential primaries unfold in perhaps the most polarized political climate since the Civil War, there is one thing that unites both parties — shock at the success of Donald Trump’s candidacy and his ability to galvanize a group of Americans despite weekly gaffes and a generally “unpresidential” demeanor.

To date, the results of the primaries have defied prediction after prediction from even the most reputable pollsters and election forecasters. Though this project will not examine Trump specifically, we think it is a step towards understanding the relationship between the political climate and the current American electorate, and perhaps also a way to understand how an apparently bigoted, alarmingly unpredictable, and unqualified candidate has amassed such strong support.

III. RELATED WORK

We have drawn ideas and inspiration from several related works, summarized below.

Psychological aspects of natural language use: Our words, our selves [1] discusses how text analysis can reveal insights into someone’s personality, psychological state, and social and situational fluctuations. The authors of the paper also examine particles — pronouns, articles, prepositions, etc. — which is interesting because many similar text analysis projects deliberately remove this group of words prior to analysis. The idea behind the paper is that word choice can hint at the speaker’s (or writer’s) age, sex, and social status,

among many other characteristics. As the paper mentions, this is not a new concept; it’s been around at least since Freud made this observation in 1901. The authors examine linguistic style rather than linguistic content, as we will be doing in our project. The authors explain the subcategory of quantitative text analysis that we will be using in our project: word count strategies. These strategies are based on the assumption that a person’s word choice reveals something greater than the literal meaning of the words and separate from their context. In addition to counting unique words, it may be useful to count a “type” of word, such as “emotion words” or “achievement-related words.” The paper reviews six widely-used methods of text analysis: The General Inquirer, Analyzing Emotion-Abstraction Patterns: TAS/C, Weintraub’s Analysis of Verbal Behavior, Analyzing Verbal Tone with DICTION, Linguistic Inquiry and Word Count, and Biber: Factor Analyzing the English Language. Next the paper discusses how to use word count strategies to draw conclusions about demographic variables such as age and sex, then personality measures, mental health and psychopathology, and physical health and health behavior. There are also techniques to guess whether the speaker is in a formal or informal setting, whether he or she is being deceitful or honest, and whether he or she is particularly stressed or experiencing trauma. The authors suggest that pronouns may provide insight into a person’s level of social integration as well as self-focus, and conclude that future research should explore the nature of pronouns and other articles in much greater detail.

In contrast, *Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts* [2] focuses on analysis of linguistic content rather than linguistic style. This paper centers on the idea that language serves as a medium for politics, and consequently is also a medium for political conflict. Candidates and elected representatives write and debate legislation, and use language to express their platforms and positions. Because language is central to understanding politics, we run into the problem of volume; political texts are too numerous and too lengthy to all be read by scholars or researchers. Automated methods provide a solution to this problem.

The authors emphasize that automated methods do not serve as a true alternative to careful reading, and suggest that these methods are “best thought of as amplifying and augmenting careful reading and thoughtful analysis.” As such, the paper presents and explains four principles of quantitative text analysis: (1) All quantitative models of language are wrong—but some are useful. (2) Quantitative methods for text amplify resources and augment humans. (3) There is no globally best method for automated text analysis. (4) Validate, Validate, Validate.

Validate. Following the descriptions of these principles, the authors discuss acquiring text data, data cleaning, stopwords, and several algorithms and examples of text content analysis.

Another paper, *The impact of linguistic strategies in political debate* [3], looks at the different devices used in political debates in order to persuade the audience. The three devices that are studied here are: Political rhetoric, Politeness strategies, and propaganda language. These three devices are identified and assessed in the 2004 presidential debate involving John Kerry and George W. Bush.

The study found that though all three of the devices were present, they were rarely used alone. Most times when it was clear that a certain device was being used, another device was also being used. This is most evident in the slogans that found a strong foothold in the debate. Such slogans used opposites as contrastive pairs in order to push each other on the issues. The final argument of this paper was a nice touch. Due to the amount of political rhetoric used in the debate the author pleads with the readership to make sure to think for themselves and not take politicians words at face value.

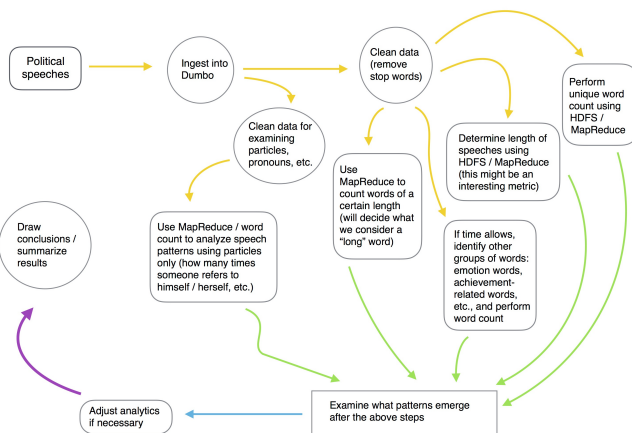
The final paper we read during our research, *Textual, intertextual and rhetorical features in political discourse: the case of President Obama in Europe* [4], explores the rhetorical, intertextual and textual properties of Barack Obama's speech given in 2009 in Strasburg. This speech had particular importance since he was urging Europe to help with the Afghan war which they — at the time — did not agree with.

There has long been an understanding that political discourse has some features that must be constant which allow the audience to recognize the speech as one that is political. These features are examined in the speech Obama gave in 2009. Being one of the best orators of our time, Obama has a wide arsenal of political idioms and analogies. By examining his extensive speech history, we can find patterns in his language that stay true to the form of political discourse while giving unique style and creativity that makes his speeches distinctive.

The paper found that the 2009 speech broke all the traditional rules predicted by Obama's earlier speeches and that, in fact, the sullen nature of the speech revealed the hidden be-guiling amongst the words themselves. It is with similar analysis that our project will reveal hidden inflections amongst our many presidents.

IV.

DESIGN

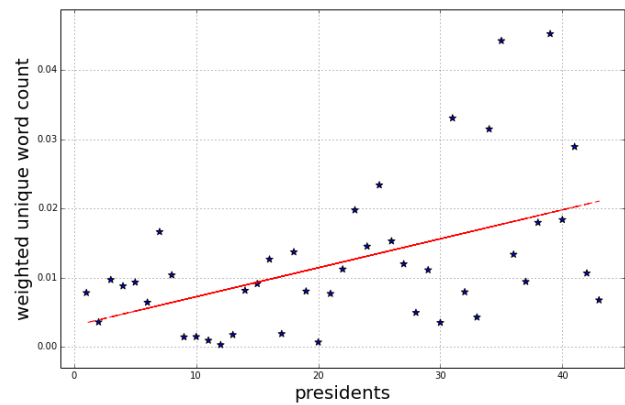


V.

RESULTS

Getting all of our data in a sensible format for our particular analytics was a much longer and more arduous process than we anticipated. We used scrapers to pull speeches from the Miller Center Foundation (<http://millercenter.org/president/speeches>) and from The American Presidency Project (<http://www.presidency.ucsb.edu/>). After a few false starts, we decided to format data as JSON objects as we scraped.

In our initial R&D, it seemed that our initial hypothesis — that over the last century, the vernacular used to speak to the electorate has been drastically simplified — was correct. But once we processed a larger amount of data, an even more interesting trend emerged: instead of the consistent downward trend in unique word count that we expected, we saw an increase in weighted word count up until Reagan, and then weighted word count began to fall.

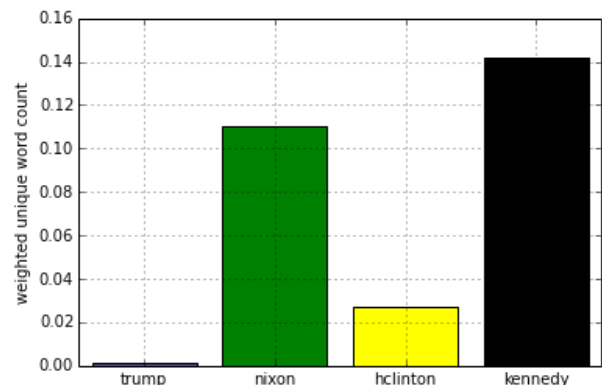


The above image shows a time-series analysis of presidents' weighted unique word count.

We had two theories about factors that might contribute to these results.

- The practice of using speech writers
- The political change that was brought on by the 1960 election and the effects of the mass media

We did a small case study to investigate the latter theory.



In the preceding graph, we compare and contrast the frontrunners in the 1960 election and those in the 2016 election. These results are more similar to the trend we expected to see.

VI. FUTURE WORK

During the planning of this project, we had all kinds of ideas about different analytics to run. We were inspired by the papers we read to perform and compare counts of different groups of words — for example, counting self-referential words per president, or counting emotion-related or achievement-related words. In the end, we didn't have time to expand our project into those areas, but we suspect there are interesting discoveries to be made there.

Campaign-specific language is another territory we think might yield interesting numbers. Future work could involve performing similar analytics on aggregated data from Twitter over the last several campaigns — specifically, we would be interested to see how Donald Trump's unique word count compares to other candidates, and how a count of anger-related and hatred-related words would look over the span of his campaign.

VII. CONCLUSION

The premise of this project was that there has been a drastic decline in the American vocabulary over the last 100 years. While this assertion was not supported by our analytics, we did discover some interesting trends in the American political vocabulary.

The data does support, however, the decline of the American political vernacular over the last fifty years, due to a spike in weighted unique word count around the Reagan-era presidencies. The reasons for this spike remain unknown. We leave this for future research.

ACKNOWLEDGMENT

The authors of this paper would like to thank the following people and institutions for either inspiring or assisting in the development of this project and paper:

Suzanne K McIntosh, Miller Center, UCSB American Presidents Project, Donald J. Trump.

REFERENCES

1. J.W. Pennebaker, M.R. Mehl, K.G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. Department of Psychology, University of Texas at Austin, 2003.
2. J. Grimmer, B.M. Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, January 2013.
3. M.A. Moskow. The impact of linguistic strategies in political debate: A linguistic discourse analysis of George W Bush and John Kerry's first presidential debate in 2004. *Institutional för Individ och Samhälle*, May 2005.
4. H.G. Conchi. Textual, intertextual and rhetorical features in political discourse: the case of President Obama in Europe. *Revista de Lingüística y Lenguas Aplicadas*, [S.l.], v. 8, p. 59-75, jun. 2013.