

# Comparing and Contrasting the Content of US Presidential Inaugural Addresses and UK Political Party Manifestos from World War II to Present Day

Ziv Schwartz  
zs1349@nyu.edu

May 8, 2020

## 1 Introduction

The purpose of this paper is to compare and contrast how prominent political figures and parties from the United States and the United Kingdom use official addresses and manifestos. This paper will focus on the content of presidential inaugural addresses from the United States and political manifestos from the United Kingdom's two largest parties: the Conservative and Labour Parties. It will be interesting to analyze the content of these documents both over time and how they differ between countries and parties. The goal of this analysis will be to elucidate if historical and political circumstances impact these documents. Different metrics will be used to help understand how these official documents have adapted over time and how leaders and governments from two prominent world leaders compare. The analysis will focus on documents dating back to the end of World War II all the way through to the present day. In what ways were the responses to the end of World War II different between the United States and the United Kingdom? How will the structure of these official documents change over time and how will two major political powers compare in how they address the evolving world?

In order to gain a better understanding, the following describes in more detail the two types of documents that will be used for analysis in this paper. First, inaugural addresses

are speeches that newly elected presidents give on the day in which they are inaugurated into office: January 20. Presidents elected to two terms therefore give two inaugural addresses. The inaugural address is most often used as a platform for a president to present their ideas and goals for their administration and the nation. The second document is manifestos. The manifestos, or party platforms, are a way in which the political parties of the United Kingdom pitch their voters. More formally, manifestos are a published set of motives, intentions and goals for the political parties. Manifestos tend to vary in length. For this analysis, the paper will focus on the two major political parties in the United Kingdom: Conservative and Labour. The key distinction between these two types of documents is that inaugural addresses are given after an official is elected whereas the manifestos serve as a platform for constituents to use when deciding how to vote in the coming election. This is an important distinction to keep in mind during the analysis and discussion of results attributed to the different documents.

## **2 Literature Review**

There seems to be more reports discussing inaugural addresses than political manifestos. This type of analysis is not necessarily novel in that many have compared the inaugural addresses of different presidents looking to compare and contrast their styles and content of speeches. Additionally, there are underlying assumptions that can be made regarding the correlation of these addresses and the major events affecting life domestically in the United States. Previous works relating to the analysis done by this paper include the following article [Edwards 2020] which analyzes not only the type of content of the inaugural addresses but also the unique characteristics of these addresses such as tie color, number of pride poses, and which president smiled the most times during their address. The more text as data analysis involved differentiating which presidents used more emotional, logical, and individualistic language, to name a few. On a more academic note, the following paper

[Kubát and Čech 2016] provided a research driven look into the “Quantitative Analysis of US Presidential Inaugural Addresses.” These researchers looked at “vocabulary richness, thematic concentration and text activity” as metrics to differentiate between the different speeches given by presidents.

What separates this current paper from the previous research done is the intracountry analysis, coupling the inaugural addresses with UK party manifestos, and trying to gain a better understanding of how different leaders and political parties discuss major events. Additionally, the distinction between the two data sources, one post-election and the other pre-election, allows for another layer of analysis in seeing how documents are structured when the election is already won versus when the document needs to be used in order to persuade voters in an upcoming election.

### 3 Theory and Hypotheses

In terms of theory and hypothesis, one that is worth exploring is using the party of the elected president to determine the similarity between the inaugural address given to the similar party manifesto of the United Kingdom. More specifically defined, it will be interesting to determine if in the years when, for example, republican presidents are elected will their inaugural addresses be more similar to the Conservative Party manifestos. Or in years that democratic presidents were elected are their speeches more closely related to the manifestos of the Labour Party. While the party platforms are not entirely the same in the United States and the United Kingdom, each generally share a majority of either right or left wing ideologies, respectively.

### 4 Data and Methods

The data used in this paper is readily available in `quanteda` [Benoit et al. 2018]. For the inaugural addresses, the data can be found at `data_corpus_inaugural` and this paper will

focus on the addresses between 1945 and 2017. For the manifestos, the data can be found at `data_corpus_ukmanifestos` and the available manifestos in `quanteda` date from 1945 to 2005. The paper will not be able to provide distinct intracountry analysis for the most recent years (2005 to 2017) however the analysis during these years can focus solely on modern inaugural addresses, with a deeper United States and United Kingdom exploration for the bulk of the overlapping years. For preprocessing, each of the documents will be stripped of punctuation, lower-cased and removed of stopwords. Words will not be stemmed and kept in their original lemma form.

The first method used when comparing the documents is looking at how the length of each document changes over time. To do so, a document-feature matrix (dfm) is constructed around each inaugural address and party manifesto. For each document, two dfm's are built, one containing stopwords and the other without. The number of features in each is then be recorded.

Next, document entropy is calculated. Entropy is an important feature in taking the analysis further when trying to establish feature selection given the maximum entropy in order to build a classifier. Maximum Entropy models differ from other classification problems because they assume no knowledge on prior distributions and hold no assumption that documents are independent, as is the case in Naive Bayes. Seeing the entropy across different time periods and party affiliations can help explain certain assumptions that are not obvious given the particular document.

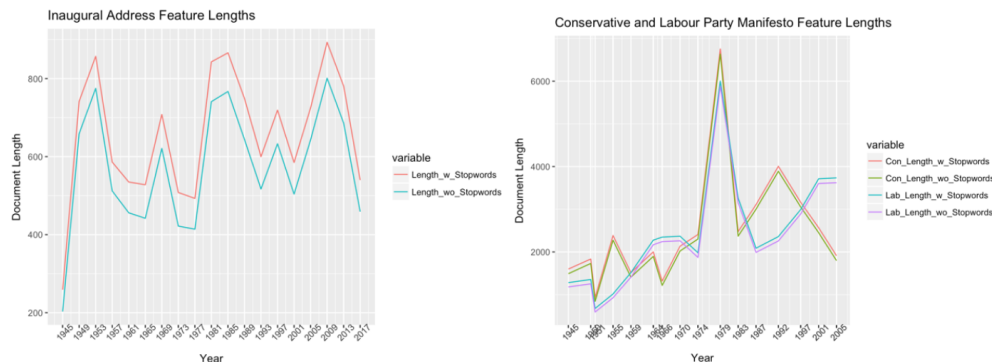
The next methodology for analyzing the differences between the different inaugural addresses and the manifestos will be to look at the Type-to-Token ratios (TTR) and Moving Average Type-to-Token ratios (MATTR) to determine the lexical complexity of the documents. Lexical complexity refers to the quantity of unique words present. It is an important area of analysis because it can lead to correlations about a president or political party's vocabulary and intellectual association. Some argue that those with larger vocabulary display higher levels of intellect, so seeing the prevalence of TTR and MATTR can validate

these hypotheses. Mathematically defined, TTR is the number of unique words ( $V$ ) over the total number of words ( $N$ ) present in the document, resulting in a number between 0 and 1. MATTR is similar in calculation, except it is calculated over a moving window of tokens from the initial to the final token, calculating TTR for each window and taking the average. Larger TTR and MATTR values indicate greater lexical complexity. The size of the token window can be tuned to better fit the lexical complexity in question. Another metric used to analyze lexical complexity is Yule's Characteristic  $K$ . Yule's Characteristic is a more complex metric that tends to be more reliable for longer and independent documents.  $K$  can better differentiate lexical complexity given a document's vocabulary being concentrated among a set of words.

Finally, looking at the cosine similarity between each address and the conservative and labour manifestos can provide a better understanding of how inaugural addresses and political party manifestos are related between the United States and the United Kingdom. A hypothesis this paper is trying to answer is if a republican or democratic president give inaugural addresses that are more similar to conservative or labour manifestos, respectively. To do this, the cosine similarity between each inaugural address and the two party manifestos is calculated for the years where there is data overlap (1945 to 2005). Not all documents are released the same year, yet the analysis allows for a few year buffer to gain a better understanding of similarity.

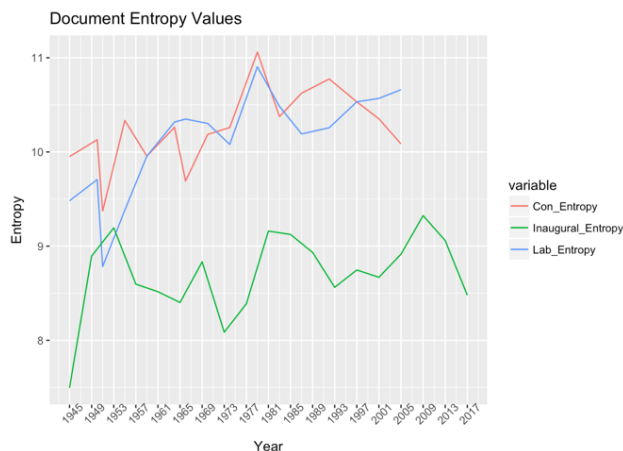
## 5 Results

### 5.1 Document Length



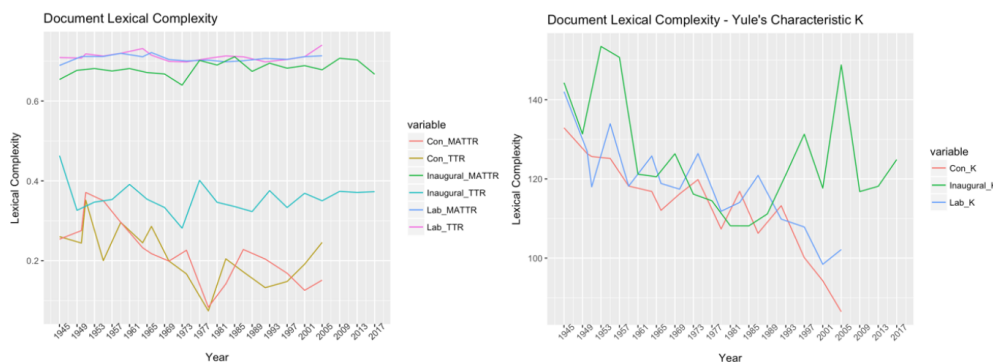
The two stacked lines specify the length of the documents with and without stopwords. The first obvious difference that can be seen between the two types of documents is that the United Kingdom manifestos are much longer in terms of feature length, being consistently greater than 1000, while inaugural addresses are not. Additionally, both the Conservative and Labour Party manifestos tend to generally be increasing in length over time, whereas the length of inaugural addresses tends to oscillate. The most striking findings from this is that the inaugural address delivered right after World War II in 1945 is the shortest of the last 19 addresses. Also, there is a massive spike in manifesto length in 1979 with both Conservative and Labour Party manifestos being longer than 6000. There is also a large disparity between conservative and labour manifestos from 2005, with the labour manifesto being almost 2000 words longer.

## 5.2 Document Entropy



Next, we look at how document entropy has changed over time and how the different documents compare. The most striking difference is that the entropy of inaugural addresses is significantly lower than both of the manifestos. Similar to feature length, the entropy of the 1945 inaugural address is the lowest and the values tend to be right around 9, while manifesto entropy is about 1.5 points higher, oscillating between 10 and 11.

## 5.3 Document Lexical Complexity



To assess lexical complexity, the following metrics: TTR, MATTR, and K, are displayed above. The MATTR is consistently higher than the TTR for all of the documents and tends to be roughly consistent from 1945 to present day. An interesting thing to note is that the TTR of the inaugural addresses is higher than the manifestos whereas the MATTR is the opposite, with the manifesto values being greater. In the simple metric for vocabulary

richness (TTR), inaugural addresses are more rich yet in the case of the moving average (MATTR), the manifestos are more rich. Next, the Yule Characteristic K is plotted over time to gain another metric for lexical complexity and it seems that for this metric, the party manifestos tend to decline in lexical complexity over time. For inaugural addresses, the K value peaks in 1953 and in 2005 where both presidents were republican (Eisenhower and George W. Bush). However, in President Bush's first address (2001) the K value was much lower, so party affiliation is not necessarily the reasoning for this increase.

## 5.4 Document Cosine Similarity

Inaugural Address Year	Conservative Party Cosine Similarity	Labour Party Cosine Similarity	Inaugural Address Year	Conservative Party Cosine Similarity	Labour Party Cosine Similarity
1945	0.3058370	0.2622773	1977	0.3687939	0.3373256
1949	0.4105366	0.3822590	1981	0.4204671	0.3810272
1953	0.4013369	0.3002594	1985	0.4273255	0.3546403
1957	0.4085219	0.2709639	1989	0.3978739	0.3334937
1961	0.2631055	0.2631055	1993	0.3400792	0.3188869
1965	0.2841622	0.2970487	1997	0.4102849	0.4337078
1969	0.3236696	0.3327850	2001	0.3149600	0.2955911
1973	0.3735902	0.3692725	2005	0.2609867	0.2813040

For the years where there is an overlap, the cosine similarity was calculated between each inaugural address and both the Conservative and Labour Party manifestos. The table above shows that 75% (12 of 16) of the inaugural address are more similar to the Conservative Party manifestos. In most years, the difference in similarity is not too significant so not much can be concluded for those. However, the largest difference between cosine similarities occurs in the years 1953 and 1957 with the inaugural addresses given by President Eisenhower, a conservative republican. This correlates with the Conservative Party manifestos where the election was won in the United Kingdom with Prime Ministers Churchill, Eden, and Macmillan.



## 6 Discussion

The results presented above provide a well-rounded analysis of the underlying properties of presidential inaugural addresses and Conservative and Labour Party manifestos dating back to World War II. The metrics looked at in this paper include: document feature length, entropy, lexical complexity and cosine similarity. Each of the metrics is analyzed in both a time series of document progression and an intracountry cross section.

The first major distinction is noting that the political manifestos are much longer than inaugural addresses. Each document serves a different purpose in the political world. Inaugural addresses are meant to present a concise vision for a president's administration and goals for the country. Thus a shorter length is expected as they are read aloud to the entire country. On the other hand, the manifestos are focused on attracting voters, therefore requiring a deeper dive into policies and platforms for the party. This also explains the large spike in manifesto length before the 1979 election, which would go on to be a watershed moment in United Kingdom politics and history. Margaret Thatcher won that election as the leader of the Conservative Party, making her the first female prime minister of the UK, and signaling a major shift in the political nature of the country. The Conservative Party ran on a platform of major reforms, indicated by the fact their party manifesto is longer. The Labour Party followed suit, resulting in both manifestos being over 6000 features long.

The next metric used is entropy, a measure that holds the ability to give a reading on how much information is present in a document. The inaugural addresses from 1945 to 2017 have an entropy that hovers around 9 whereas the party manifestos have greater entropy, reaching a peak in 1979 of right around 11 and generally between 10 and 11. This again follows the expectation of the assumptions that underlie the differences in the documents. The manifestos are meant to contain more information and specifically focusing on the 1979 manifestos, the peak entropy correlates nicely with the greater document lengths. For the inaugural addresses, the most surprising value is the minimum entropy of the 1945 address, immediately following the conclusion of World War II. A logical explanation for this could

be that the address was focused on one central topic alone, moving forward from the war, and thus limited in terms of entropy and information.

Following this, the lexical complexity of each document is calculated using three different metrics: TTR, MATTR, and Yule's Characteristic K. The simplest and most common metric for vocabulary richness is TTR, which calculates the ratio of unique words over total words in a document. 1979 again is a very interesting year to look at in terms of lexical complexity. Given that the manifestos from 1979 both contain the most features and highest entropy, it is not expected that the 1979 manifestos would have the lowest lexical complexity. However, given the significance of the election, the architects of these manifestos decided to focus on clear, simple and concise words in order to get their party message across and not a unique vocabulary. Looking at Yule's Characteristic K, it seems that UK manifestos decrease in lexical complexity over time, yet inaugural addresses do not follow the same trend, with peaks in 1953 and 2005.

Finally, a similarity metric is used to see how closely inaugural addresses relate to either the conservative or labour manifestos. From the table, 75% of the addresses have a higher cosine similarity to the Conservative Party manifestos. The hypothesis presented earlier speculates that republican presidents would be expected to have a higher similarity to Conservative Party manifestos and democratic presidents would have a higher similarity to the Labour Party manifestos, however this is not generally the case. This seems to only be true with the inaugural addresses of President Eisenhower in 1953 and 1957.

Future work will be to look more closely at topic modeling and trying to develop a methodology for establishing the relationship between the key topics found in these documents and domestic and international events. It is a more intensive analysis to look at topics from across a wide range of dates and mapping them correctly to major events. One way this can be done will be to bring another data source of major events, whether domestic or international, and seeing if the topics extracted from the addresses of the most prominent political figures correlate to these events.

## References

- [Ben+18] Kenneth Benoit et al. “quanteda: An R package for the quantitative analysis of textual data”. In: *Journal of Open Source Software* 3.30 (2018), p. 774. DOI: 10.21105/joss.00774. URL: <https://quanteda.io>.
- [Edw20] Vanessa Van Edwards. *Trump to Truman: An Analysis of 20 Presidential Inaugural Speeches*. Apr. 2020. URL: <https://www.scienceofpeople.com/inauguration-speech/>.
- [KČ16] Miroslav Kubát and Radek Čech. “Quantitative Analysis of US Presidential Inaugural Addresses”. In: *Glottometrics* 15 (Jan. 2016), pp. 14–27.