

## Article

# Fake News and Propaganda: Trump's Democratic America and Hitler's National Socialist (Nazi) Germany

David E. Allen <sup>1,2,3</sup>  and Michael McAleer <sup>2,4,5,6,7,8,\*</sup> 

<sup>1</sup> Department of Finance, School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia; profallen2007@gmail.com

<sup>2</sup> Department of Finance, College of Management, Asia University, Wufeng 41354, Taiwan

<sup>3</sup> School of Business and Law, Edith Cowan University, Joondalup, WA 6027, Australia

<sup>4</sup> Discipline of Business Analytics, University of Sydney Business School, NSW 2006, Australia

<sup>5</sup> Econometric Institute, Erasmus School of Economics, Erasmus University, 3062 Rotterdam, The Netherlands

<sup>6</sup> Department of Economic Analysis and ICAE, Complutense University of Madrid, 28040 Madrid, Spain

<sup>7</sup> Department of Mathematics and Statistics, University of Canterbury, Christchurch 8041, New Zealand

<sup>8</sup> Institute of Advanced Sciences, Yokohama National University, Yokohama, Kanagawa 240-8501, Japan

\* Correspondence: michael.mcaleer@gmail.com

Received: 18 May 2019; Accepted: 21 August 2019; Published: 21 September 2019



**Abstract:** This paper features an analysis of President Trump's two State of the Union addresses, which are analysed by means of various data mining techniques, including sentiment analysis. The intention is to explore the contents and sentiments of the messages contained, the degree to which they differ, and their potential implications for the national mood and state of the economy. We also apply Zipf and Mandelbrot's power law to assess the degree to which they differ from common language patterns. To provide a contrast and some parallel context, analyses are also undertaken of President Obama's last State of the Union address and Hitler's 1933 Berlin Proclamation. The structure of these four political addresses is remarkably similar. The three US Presidential speeches are more positive emotionally than is Hitler's relatively shorter address, which is characterised by a prevalence of negative emotions. Hitler's speech deviates the most from common speech, but all three appear to target their audiences by use of non-complex speech. However, it should be said that the economic circumstances in contemporary America and Germany in the 1930s are vastly different.

**Keywords:** text mining; sentiment analysis; word cloud; emotional valence

**JEL Classification:** C19; C65; D79

## 1. Introduction

President Trump continues to attract controversy in the media and in political commentary, partly because of his attitude to “fake news”, combined with his own lavish use of his Twitter account and lack of attention to the verification of some of his more extreme pronouncements. In 2018, the President used Twitter to announce the “winners” of his “fake news” awards, most frequently naming the New York Times and CNN for a series of perceived transgressions which varied from minor errors by journalists on social media to news reports that later invited corrections.

Given his predilection for criticising the media, the authors have previously analysed his pronouncements on climate change [1], on nuclear weapons and [2], and contrasted his first State of the Union Address (SOU) with the previous one by President Obama [3].

Given the controversy about the timing and delivery of his most recent SOU address, the authors thought it might be of interest to subject both of his SOU addresses to textual analysis using data mining techniques, so as to explore whether his political addresses are typical or whether they deviate markedly from those of other political leaders, specifically, Obama and Hitler, the latter being selected as an extreme benchmark. The null hypothesis is that political speeches are essentially similar.

We decided to analyse both Trump's 2018 State of the Union Address (SOU1), and 2019 address (SOU2) to assess whether there had been any change in the structure and emotional tenor of his two addresses in response to changing political and economic circumstances, at the end of the second year of his term in office. To provide a contrast, one contemporary and another more historically extreme, we also analyse President Obama's last SOU and Hitler's 1933 Berlin Proclamation.

The contents of these speeches are analysed using a variety of R packages, including several in data mining: "tm" a text mining package, created by Feinerer and Hornik [4]. We also used "syuzhet", a sentiment extraction tool, originally developed in the NLP group at Stanford University, and then incorporated into an R package by Jockers [5], and "wordcloud" by Fellows [6].

Data mining refers to the process of analysing datasets to reveal patterns, and usually involves methods that are drawn from statistics, machine learning, and database systems. Text data mining similarly involves the analysis of patterns in text data. Sentiment analysis is concerned with the emotional context of a text, and seeks to infer whether a section of text is positive or negative, or the nature of the emotions involved. There is a variety of methods and dictionaries that exist for undertaking sentiment analysis of a piece of text.

Although sentiment is often framed in terms of being a binary distinction (positive versus negative), it can also be analysed in a more nuanced manner. We decided to apply the R package "syuzhet", which distinguishes between eight different emotions, namely trust, anticipation, fear, joy, anger, sadness, disgust and surprise. There are many different forms of sentiment analyses, but most use the same basic approach. They begin by constructing a list of words or dictionary associated with different emotions, count the number of positive and negative words in a given text, and then analyse the mix of positive and negative words to assess the general emotional tenor of the text.

Clearly, there are considerable limitations to the basic approach adopted in the paper. Pröllochs et al. [7] discussed the difficulties in processing negations, which invert the meanings of words and sentences. Equally problematic are sarcasm, backhanded compliments, and inflammatory gibberish, such as "Pocahontas" and "Crooked Hillary", in the context of President Trump's tweets. Nevertheless, sentiment analysis can reveal the general emotional direction of a piece of text, and machine-based learning systems are well-established methods for the sifting and interpretation of digital information. This tool has numerous applications in, for example, financial markets.

We can now apply machine learning techniques to news feeds to determine what average opinion is. For example, the Thomson Reuters News Analytics (TRNA) series could be termed news sentiment, and is produced by the application of machine learning techniques to news items. The TRNA system can scan and analyse stories on thousands of companies in real time, and translate the results into a series that can be used to help model and inform quantitative trading strategies. RavenPack is another example of a commercial news analytics product that has applications to financial markets. There is now considerable evidence about the commercial relevance of financial news analysed using machine learning methods.

Allen, McAleer and Singh [8,9] analysed the economic impact of the TRNA sentiment series. The first of these papers examines the influence of the Sentiment measure as a factor in pricing DJIA constituent company stocks in a Capital Asset Pricing Model (CAPM) context. The second uses these real time scores, aggregated into a DJIA market sentiment score, to analyse the relationship between financial news sentiment scores and the DJIA return series, using entropy-based measures. Both studies find that the sentiment scores have a significant information component which, in the former, is priced as a factor in an asset pricing context.

Allen, McAleer and Singh [10] used the Thomson Reuters News Analytics (TRNA) dataset to construct a series of daily sentiment scores for Dow Jones Industrial Average (DJIA) stock index constituents. The authors used these daily DJIA market sentiment scores to study the influence of financial news sentiment scores on the stock returns of these constituents using a multi-factor model. They augmented the Fama–French three-factor model with the day’s sentiment score along with 20 lagged scores to evaluate the additional effects of financial news sentiment on stock prices in the context of this model. Estimation is based on Ordinary Least Squares (OLS) and Quantile Regression (QR) to analyse the effects around the tails of the returns distribution. The results suggest that, even when market factors are taken into account, sentiment scores have a significant effect on Dow Jones constituent returns, and that lagged daily sentiment scores are also often significant.

Other research on this topic argues that news items from different sources influence investor sentiment, which feeds into asset prices, asset price volatility and risk (see, among others, Tetlock [11] Tetlock, Macskassy and Saar-Tsechansky [12] (2008), Da, Engleberg and Gao, [13], Barber and Odean [14], diBartolomeo and Warrick [15], Mitra, Mitra and diBartolomeo [16], and Dzielinski, Rieger and Talpsepp [17]. The diversification benefits of the information impounded in news sentiment scores provided by RavenPack were demonstrated by Cahan, Jussa and Luo [18], and Hafez and Xie [19], who examined the benefits in the context of popular asset pricing models.

Several papers provide surveys of this burgeoning literature. Kearney and Lui [20] concentrated on sentiment analysis and provided an analysis of methods and the related literature. Loughran and McDonald [21] provided a survey of the accounting, finance, and economics literature on textual analysis, plus a description of some of its methods, together with potential pitfalls in its application.

In the current paper, the focus is on the actual content of President Trump’s 2018 SOU1, and his subsequent 2019 SOU2 address. The intention is to explore whether there are any systematic differences in the sentiments of these two SOUs, and whether there is any evidence of a tendency by President Trump to generate a “positive” spin for the benefit of his voter base. A contrast is provided by parallel analyses of President Obama’s last SOU and Hitler’s 1933 Berlin Proclamation.

Could President Trump’s addresses be fairly described as constituting “propaganda”? This has been defined as being the presentation of information, ideas, opinions, or images, which may only present one part of an argument, and which are broadcast, published, or in some other way spread with the intention of influencing people’s opinions. Sentiment analysis will not give a clear answer as to whether content represents propaganda per se, but it will give an indication as to the emotional tenor of a text or speech. It will reveal correlations between the use of words, changes in sentiment, and any patterns revealed through time in the presentation of a speech.

An alternative approach to the analysis of language as a whole, was first suggested by Zipf (1932, p. 1) [22], who applied a concept of relative frequency which suggested that: “the accent or degree of conspicuousness of any word, syllable, or sound is inversely proportionate to the relative frequency of that word, syllable, or sound, among its fellow words, syllables, or sounds in the stream of spoken language. As any element’s usage becomes more frequent, its form tends to become less accented, or more easily pronounceable, and vice versa. He analysed whether the modern vernacular of Beijing, China, was consistent with Indo-European tongues in substantiating his “Principle of Relative Frequency”.

Zipf [22] suggested that there are four important characteristics that are recognisable in words: The first is meaning, an elusive concept which is difficult to describe. The second he described as being “quality”, by which he meant positive or negative qualities. These are the subject of sentiment analysis in the current paper. The third he described as being “emotional intensity”, which could also be related to the degree to which sentiment is espoused. The fourth he described as being “order”, which is related to semantic change and the relative frequency of use of different words. Order is also related to the probability of occurrence of different words. Zipf suggested that the formula for abbreviation is  $ab^2 = k$ .

Mandelbrot [23] expanded on this approach, refining Zipf's theory by suggesting that human languages evolved over time to optimise the capacity to convey information from the sender to receiver. He couched his analysis in terms of Shannon's [24] "information theory". Mandelbrot suggested that, as a first approximation,  $i(r, k)/k$ , which he defines as the relative number of repetitions of the word  $W(r)$  in a sample of length  $k$ , is inversely proportional to 10 times  $r$ ,  $i(r, k)/k = 1/10r$ .

Shannon ([24], p. 6) suggested that it is possible to use artificial languages to approximate natural languages. The zero-order approximation is to choose all letters with the same probability and independently. The first-order approximation is to choose each letter independently but with the same probability of occurrence as would apply in the relevant natural language. In a third-order approximation, a trigram structure is adopted with the probability of each letter dependent on the preceding two letters.

Shannon [24] suggested that we let  $p(B_i)$  be the probability  $B_i$  of a sequence of symbols from a source text. Let:

$$G_N = -\frac{1}{N} \sum_i p(B_i) \log p(B_i), \quad (1)$$

where the sum is over all sequences  $B_i$  containing  $N$  symbols. This suggests that  $G_N$  is a monotonically decreasing function of  $N$ , and that:

$$\lim_{N \rightarrow \infty} G_N = H.$$

Shannon lets  $p(B_i, S_j)$  be the probability of sequence  $B_i$  being followed by symbol  $S_j$  and  $p_{B_i} S_j = p(B_i, S_j)/p(B_i)$  be the conditional probability of  $S_j$  after  $B_i$ . Then, let:

$$F_N = -\sum p(B_i, S_j) \log p_{B_i}(S_j), \quad (2)$$

where the summation is over all blocks  $B_i$  of  $N - 1$  symbols and over all symbols  $S_j$ ; then,  $F_N$  is a monotonically decreasing function of  $N$ :

$$F_N = NG_N - (N - 1)G_{N-1},$$

$$G_N = \frac{1}{N} \sum_{N=1}^N F_N,$$

$$F_N \leq G_N,$$

and,  $\lim_{N \rightarrow \infty} F_N = H$ .

Shannon [24] stated that  $F_N$  is the entropy of the  $N$ th-order approximation to the source of the type discussed above. Mandelbrot [23] suggested that his derivation of the law of word frequencies was characterised by maximising Shannon's "quantity of information" under certain constraints.

We use some of these concepts in the subsequent analysis of the political addresses featured in this paper to explore how far they deviate from standard patterns of language. The most recent comprehensive use of this type of analysis is that of Ficcadenti et al. [25], which also features a lengthy review of the relevant literature. However, there is no sentiment analysis of Presidential speeches in their study.

The remainder of the paper is divided into four sections. An explanation of the research method is given in Section 2. Section 3 presents the results. Section 4 provides some concluding comments.

## 2. Research Method

The analysis features the use of a number of R libraries which facilitate data mining and sentiment analysis, namely word cloud, tm and syuzhet, plus a variety of graphics packages. The R package tm has a focus on extensibility based on generic functions and object-oriented inheritance, and provides a basic infrastructure required to organise, transform, and analyse textual data. The basic document is

imported into a “corpus”, which is then transformed into a suitable form for analysis using stemming, stopword removal, and so on. Then, we can create a term-document matrix from a corpus which can be used for analysis.

Once we have the text in matrix form, a huge amount of R functions (e.g., clustering, classifications, among others) can be applied. We can explore the associations of words, correlations, and so forth, and screen the text for frequently occurring words. The analysis can be used to create a word cloud of the most frequently used words. Feinerer and Hornik [4] provided an introduction to the package.

The R package wordcloud by Fellows [6] provides functionality to create word clouds, visualise differences and similarity between documents, and avoid over-plotting in scatter plots with text. We use the R package “syuzhet” for sentiment analysis. The package comes with four sentiment dictionaries, and provides a method for accessing the robust, but computationally expensive, sentiment extraction tool developed in the NLP group at Stanford University. We transform the text in character vectors. Once we have the vector, we can select which of the four sentiment extraction methods available in “syuzhet” to employ. We use the default syuzet lexicon, which was developed in the Nebraska Literary Lab under the direction of Jockers [5].

The name “Syuzhet” comes from the Russian Formalists Shklovsky [26] and Propp [27], who divided narrative into two components, the “fabula” and the “syuzhet”. “Syuzhet” refers to the “device” or technique of a narrative, whereas “fabula” is the chronological order of events. “Syuzhet”, therefore, is concerned with the manner in which the elements of the story (fabula) are organised (syuzhet). The R syuzhet package attempts to reveal the latent structure of narrative by means of sentiment analysis, and we can construct global measures of sentiment into eight constituent emotional categories, namely trust, anticipation, fear, joy, anger, sadness, disgust and surprise.

While these global measures of sentiment can be informative, they tell us very little in terms of how the narrative is structured and how these positive and negative sentiments are activated across the text. To explore this, we plot the values in a graph where the x-axis represents the passage of time from the beginning to the end of the text, and the y-axis measures the degrees of positive and negative sentiment.

President Trump’s first SOU in 2018 contained 5169 words and 30,308 characters, while his second SOU in 2019 contained 5493 words and 32,204 characters. Therefore, the two addresses were of similar size.

We use the R package “tm” and develop the appropriate R code to undertake the Zipf and Mandelbrot power law distribution analysis to assess the degree to which the four political addresses deviate from common language.

The limitations of the analysis should be borne in mind. The context of “natural language processing”, of which sentiment analysis is a component, is important. The use of sarcasm and other types of ironic language are inherently problematic for machines to detect, especially when viewed in isolation.

### 3. Results and Interpretation of the Analysis

#### 3.1. Sentiment Analysis

Figure 1 presents a word cloud analysis of President Trump’s two SOUs. In his first 2018 SOU, depicted in Figure 1A, the most frequently occurring word is “American”, followed by the symbol  $\$$ , which is a generic representation of different dollar amounts mentioned at various stages in his address. Other words emphasised include “will”, “year”, “one”, “tonight”, “people”, “new”, “year”, “america”, “together”, “great”, “home”, “tax”, “congress”, “families”, “countries”, “proud”, “just”, “job”, and “citizen”.

The second and most recent SOU by President Trump is shown in Figure 1B. This is dominated by the words “will”, “American”, “years”, “one”, “new”, “thank”, “americans”, “tonight”, “now”, “can”, “must”, “congress”, “border”, “last”, “time”, “also”, and “country”.

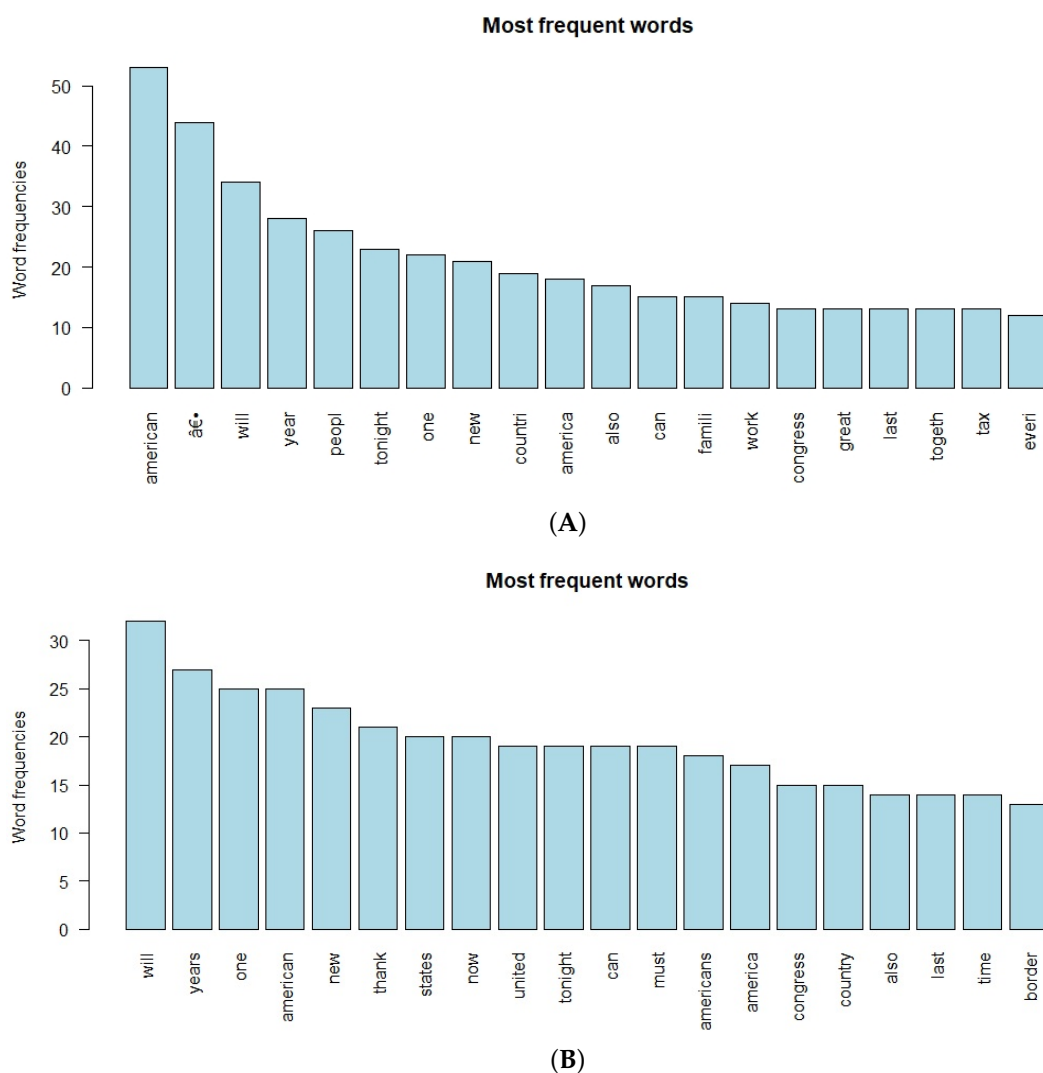






occurs over 50 times, followed by various indications of dollar amounts, “will” occurs more than thirty times, while “great”, “last”, “together” and “tax” occur around twenty times each.

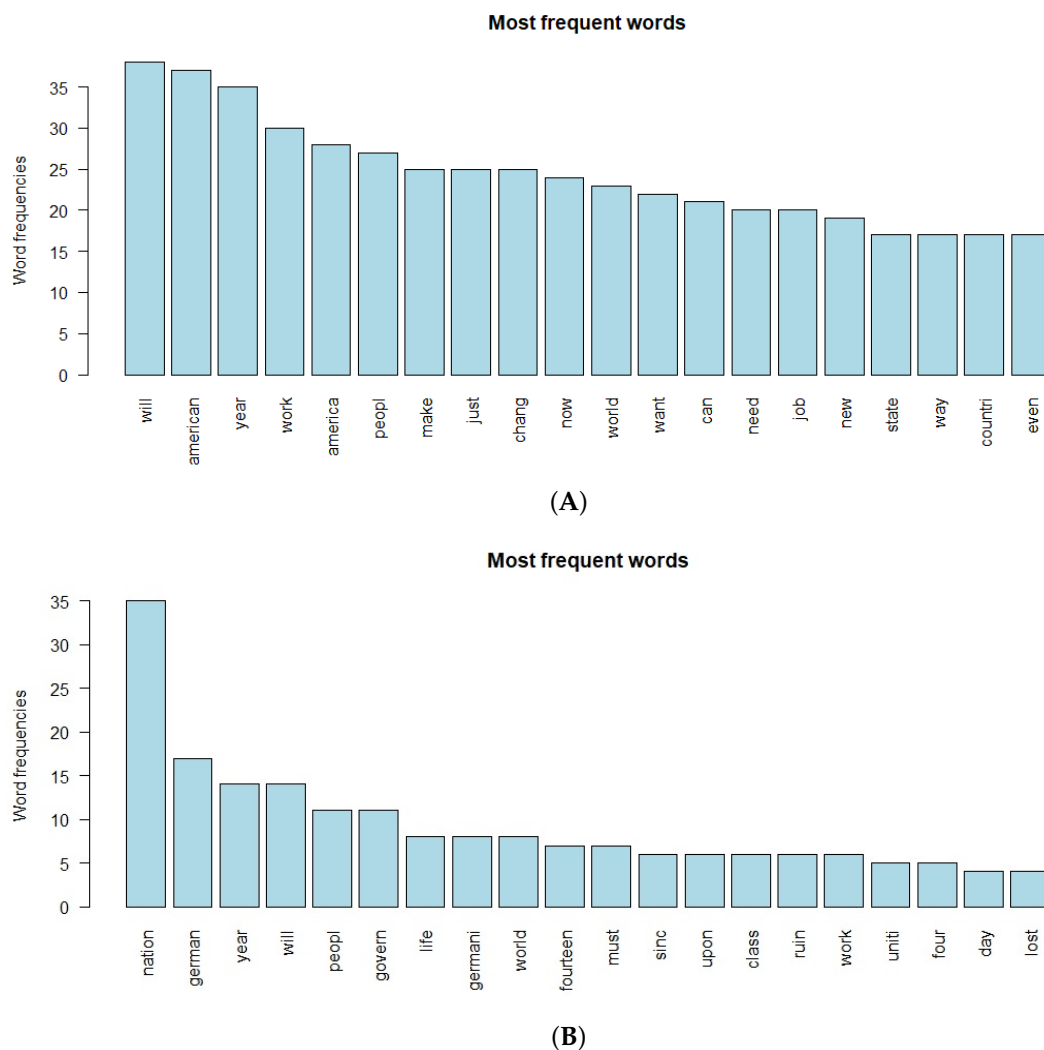
In Trump’s second SOU, depicted by the bar chart in Figure 3B, “will” becomes the most frequently occurring word, followed by “years”, “one” and “American”, but the top few words are less frequent in President Trump’s second SOU than in his first. “American” is now the fourth most frequent word rather than the first, as in the previous SOU. Perhaps surprisingly, given the political battles enveloping the topic, “border” is the twentieth most frequently used word.



**Figure 3.** Bar Plots of words used frequently in President Trump’s two SOUs: (A) President Trump SOU1; and (B) President Trump SOU2.

Figure 4 provides a similar analysis for President Obama’s last SOU and for Hitler’s 1933 Proclamation. Figure 4A reveals that the most frequently used word in President Obama’s last SOU was “will”, which occurred 38 times, closely followed by “American” 37 times, and “year” 35 times. “Work”, “America” and “people” were the next most frequently occurring words.





**Figure 4.** Bar Plots of most frequently used words in President Obama's last SOU and in Hitler's 1933 Proclamation: (A) President Obama's last SOU; and (B) Hitler's 1933 Proclamation.

Hitler's 1933 Proclamation was a much shorter speech than the SOUs just considered. However, it was relatively dominated by the word "nation", which occurred 35 times, while the next most frequently used word was "German", mentioned 17 times, while "year" and "will" occurred 14 times each.

Patriotism and nationalism appear to be frequently occurring themes in these four very different political addresses. "American" is the first and fourth most frequently occurring words in President Trump's two SOUs, and it is the second most frequently used word in President Obama's last SOU. The most frequently used word in Hitler's 1933 Proclamation was "Nation", which had double the frequency of any other words mentioned, followed by "German". There is clearly a strong nationalistic tone in his 1933 address.

The other recurrent theme in these four political speeches is the importance of intention, as captured by the use of the word "will". It is the third and first most frequently occurring word used in President Trump's two SOUs, respectively. It is the most frequent word in President Obama's last SOU and the fourth most frequently occurring word in Hitler's 1933 Proclamation.

Table 1 shows the words most highly correlated with President Trump's frequently used words in his two SOUs. "American" is the most frequently used word in his first SOU. Its use is most highly correlated with: "bridge", "gleam", "grit", "heritage", "highway", "railway", "reclaim", "waterway", "background", "color", "creed", "dreamer", "official", "religion", and "sacred".

**Table 1.** Words highly correlated with frequently used words in President Trump’s SOUs.

Trump SOU2018			Trump SOU2019		
Word	Correlated Words	Correlation	Word	Correlated Words	Correlation
American	bridge	0.34	Will	never	0.49
	gleam	0.34		Afghan	0.41
	grit	0.34		constructive	0.41
	heritage	0.34		counter terrorism	0.41
	highway	0.34		focus	0.41
	railway	0.34		groups	0.41
	reclaim	0.34		indeed	0.41
	waterway	0.34		taliban	0.41
	background	0.34		talks	0.41
	color	0.34		troop	0.41
	creed	0.34		agreement	0.38
	dreamer	0.34		achieve	0.37
	official	0.34		make	0.37
	religion	0.34		progress	0.37
	sacred	0.34		proudly	0.37
	dream	0.33		dream	0.37
	hand	0.33		holding	0.37
	land	0.31		whether	0.35
	duty	0.31		incredible	0.32
	right	0.31		back	0.51
will	arsenal	0.44	American	soldiers	0.40
	deter	0.44		astronauts	0.37
	magic	0.44		Buzz	0.37
	part	0.44		space	0.37
	someday	0.44		intellectual	0.37
	unfortunate	0.44		property	0.37
	use	0.44		Dachau	0.37
	weapon	0.44		second	0.37
	yet	0.44			
	aggression	0.40			
	moment	0.32			
	modern	0.32			

A second frequently used word is “will”, which is highly correlated with “deter”, “magic”, “part”, “someday”, “unfortunate”, “use”, “weapon”, and “yet”. The same two words are reversed in relative frequency of use in the second SOU. “Will” is most highly correlated with “never”, followed by “Afghan”, “constructive”, “counter-terrorism”, “focus”, “groups”, “indeed”, “Taliban”, “talks”, and “troop”. “American” is most highly correlated with “back” and “soldiers”.

The analysis is concerned with an examination of the extent to which political speeches by different political leaders differ. We would expect to see similarities in the two speeches by President

Trump. This includes similarities in the usage of words and correlations between pairs of words when they are made by the same politician.

Table 2 provides an analysis of the words most highly correlated with frequently used words in President Obama's last SOU and Hitler's 1933 Proclamation. The analysis of President Obama's last SOU reveals the weaknesses of a statistical analysis of individual words used as components of a particular address. The words most correlated with the word "American" were individual dollar amounts. "Will" is highly correlated with "preserve", "status-quo", and "planet".

**Table 2.** Words highly correlated with frequently used words in President Obama's last SOU and Hitler's 1933 Proclamation.

Obama SOU			Hitler 1933		
Word	Correlated Words	Correlation	Word	Correlated Words	Correlation
American	various numbers	n.a.		life	0.42
will	preserve	0.44	Nation	will	0.40
	status-quo	0.44		govern	0.37
	planet	0.30		regard	0.32
America	George Washington Carver	0.36	will	health	0.50
	Katherine Johnson	0.36		lead	0.40
	Sally Ride	0.36		nation	0.40
	unit	0.35		back	0.33
			German	assist	0.33
				work	0.34
				rescue	0.32
				support	0.32

"America" is highly correlated with individual names, the components of which the program picked up individually, and it was not until the authors analysed the original text that the analysis made sense. In the speech, President Obama stated: "Now, that spirit of discovery is in our DNA. America is Thomas Edison and the Wright Brothers and George Washington Carver. America is Grace Hopper and Katherine Johnson and Sally Ride. America is every immigrant and entrepreneur from Boston to Austin to Silicon Valley racing to shape a better future".

The analysis of Hitler's 1933 Berlin Proclamation was more revealing. "Nation", the most frequently used word, is highly correlated with "life", "will", "govern", and "regard". "Will" is highly correlated with "health", "lead", "nation", "back", and "assist". Finally, "German" is highly correlated with "work", "rescue", and "support". This supports the national rebuilding of the German economy and the promotion of employment that was part of Hitler's agenda in the early 1930s. He adopted the view that the natural unit of mankind was the Volk ("the people"), of which the German people was the greatest. He also believed that the state existed to serve the Volk. This leads to a consideration of "National Socialism" (or "Nazism").

Smith ([28], pp. 18–19) suggested that "...nationalists have a vital role to play in the construction of nations, not as culinary artists or social engineers, but as political archaeologists rediscovering and reinterpreting the communal past in order to regenerate the community. Their task is indeed selective—they forget as well as remember the past—but to succeed in their task they must meet certain criteria. Their interpretations must be consonant not only with the ideological demands of nationalism, but also with the scientific evidence, popular resonance and patterning of particular ethnohistories".

Nationalism holds that each nation should govern itself, free from outside interference (self-determination), and that the nation is the only rightful source of political power (popular sovereignty). It usually involves the maintenance of a single national identity, which would be based on shared social characteristics, such as shared history culture, language, religion, and politics. President Trump, with his slogan "MAGA" (make America great again), espouses a form of Nationalism.

President Obama's last SOU is not free of nationalistic sentiment. He stated that: "I told you earlier all the talk of America's economic decline is political hot air. Well, so is all the rhetoric you hear about our enemies getting stronger and America getting weaker. Let me tell you something. The United States of America is the most powerful nation on Earth, period. Period. It is not even close. It is not even close. We spend more on our military than the next eight nations combined."

However, as the mechanical and statistical form of textmining used in this paper, though revealing, is not suited to teasing out the nuances in meaning of different forms of nationalism, emphasis is placed on a statistical analysis of the text.

We also used the R package "syuzhet" to examine the the sentiment of each string of words or sentences. We calculated the overall score and the mean or average sentiment score. The results vary slightly, depending on which lexicon or base dictionary is used. Syuzhet incorporates four sentiment lexicons. The default "syuzhet" lexicon was developed in the University of Nebraska Literary Lab under the direction of Jockers [5], the creator of the R syuzhet package. This is the default lexicon. We also cross-checked using the nrc lexicon developed by Mohammad, who is a research scientist at the National Research Council Canada (NRC) (see: <http://saifmohammad.com>). However, the results were quantitatively similar, and hence are not reported in the paper.

The analysis tells us whether the speech has a predominantly positive or negative score in emotional tenor. In the case of President Trump's first SOU, the total score was 113.75 and the mean score was 0.02196. This positive sentiment score is consistent with Allen, McAleer and Reid [3], who reported similarly positive results for President Trump's first SOU, on the basis of an application of the R package "sentiment", which used a different lexicography. In the previous analysis, on the basis of a primary binary division into positive and negative sentiments, 60 per cent of the first SOU, in cases where sentiment could be ascribed, was recorded as being positive.

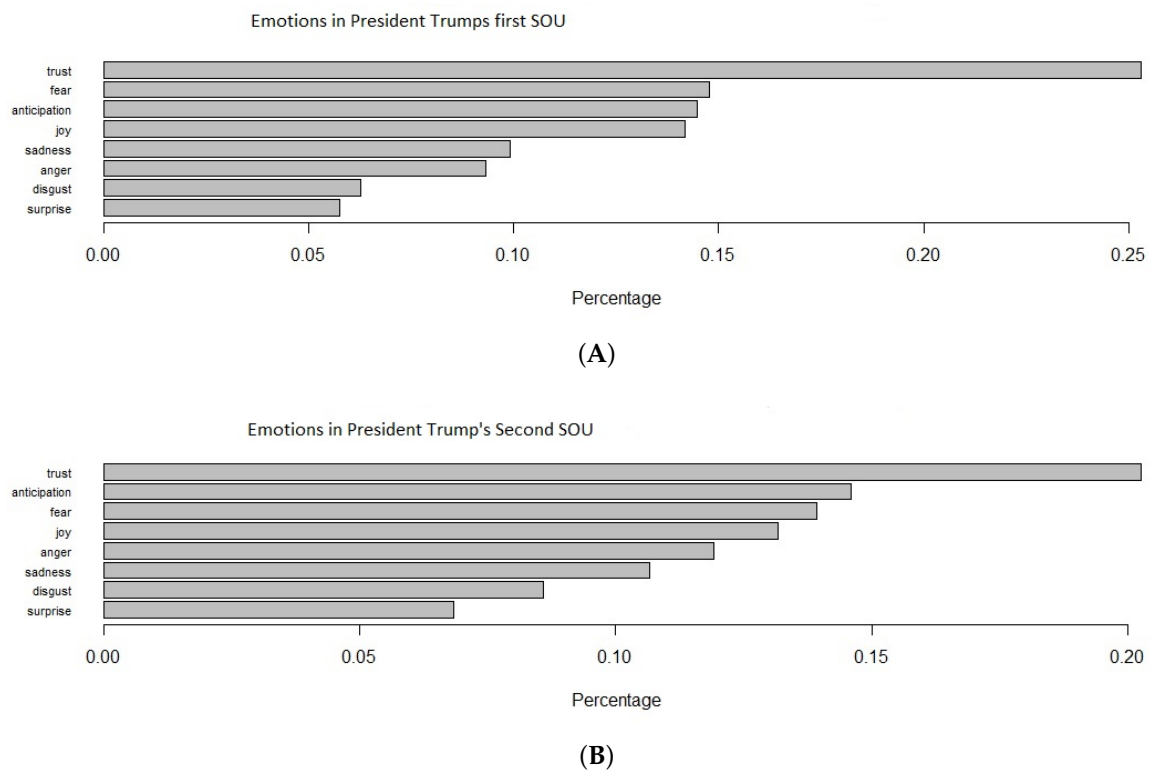
In his second SOU in 2019, the address had a total score of 139.85 and a mean score of 0.02557. His first SOU contained 5190 words and 30,271 characters, while his second SOU was slightly larger at 5442 words and 32,045 characters. President Obama's last SOU had a total score of 169.8 and a mean score of 0.02712. President Obama's last SOU was quite a large speech, containing 6233 words and 34,634 characters. In the case of Hitler's 1933 Proclamation, the sum is 8.4 and the mean is 0.0053, but Hitler's parsimonious proclamation only contained 1578 words and 9286 characters.

An interesting feature of these various speeches is the degree to which they contained predominantly positive or negative emotions. These are plotted in Figures 5 and 6. In both of President Trump's SOUs, "Trust" is the predominant emotion displayed. In all speeches, apart from President Trump's second SOU, it accounts for more than 25 per cent of the total emotional content. This is also the case in President Obama's last SOU, and in Hitler's 1933 Proclamation. In all four speeches, "Trust" dominates by a large margin in the order of 10 per cent, though it is slightly lower in President Trump's second SOU.

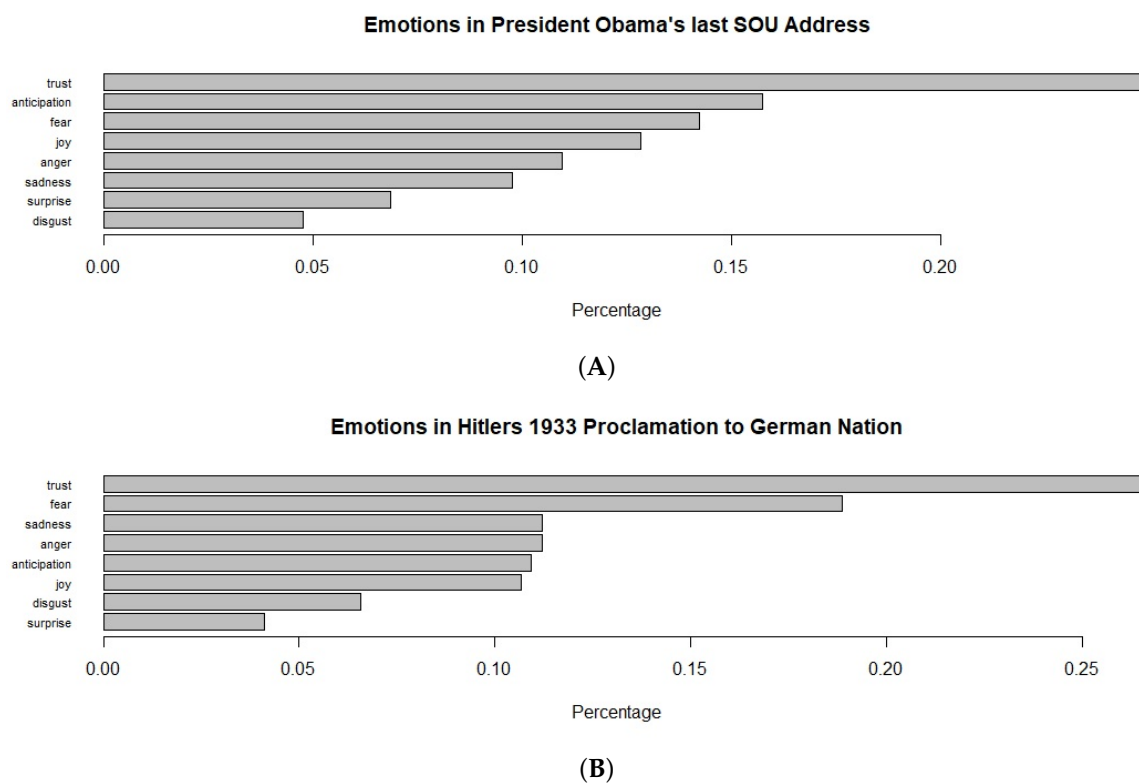
"Fear" is the second dominant emotion in Trump's SOU, and drops to third in his second SOU. "Fear" is the third emotion in President Obama's last SOU, accounting for about 14 per cent of the emotional content, but it is more prominent in Hitler's 1933 proclamation, in which it is the second ranked emotion, and accounts for about 18 per cent of the emotional content.

"Anticipation" plays a large role in President Trump's and Obama's addresses, in which it always accounts for around 15 per cent of the total emotional content; indeed, it is slightly more than 15 per cent in the case of President Obama. It is much less prominent in Hitler's Proclamation, where it is the fifth most frequently occurring emotion, accounting for about 12 per cent of the total emotional content. Indeed, a feature of Hitler's address is the predominance of negative emotions, with "fear", "sadness" and "anger" taking precedence after "trust".

In contrast, "anticipation" and "joy" are much more predominant in the two US President's SOUs, never dropping below 13 per cent in emotional content, and always ranking in the top four emotions. In Hitler's speech, "anticipation" is the fifth ranked emotion.



**Figure 5.** The Emotional Tenor of President Trump's two SOUs: (A) President Trump's First SOU; and (B) President Trump's Second SOU.

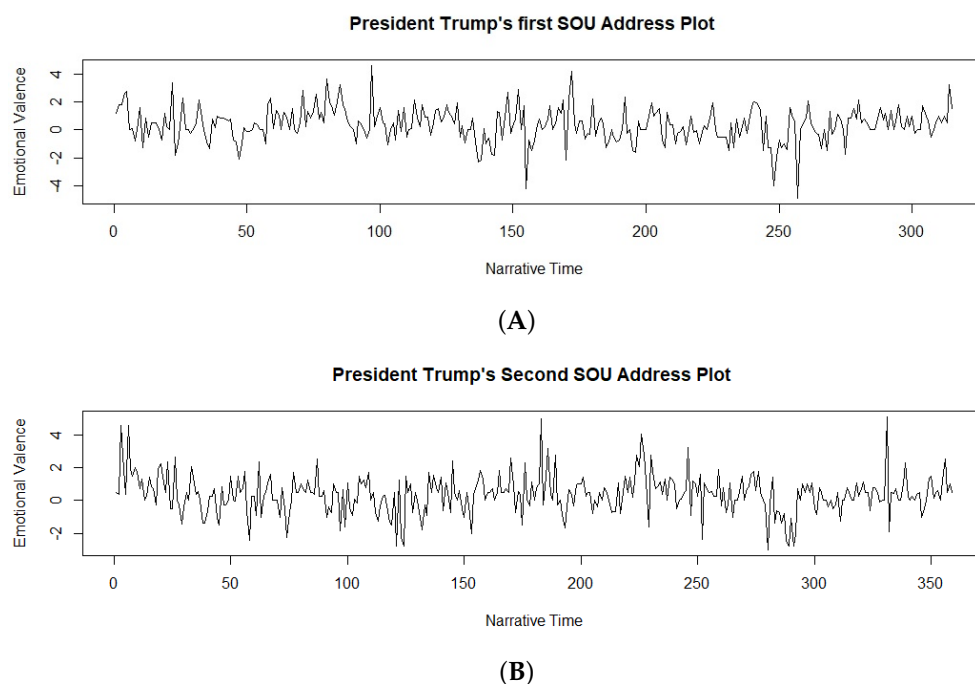


**Figure 6.** The Emotional Tenor of President Obama's last SOU and Hitler's 1933 Berlin Proclamation: (A) President Obama's last SOU; and (B) Hitler's 1933 Proclamation.

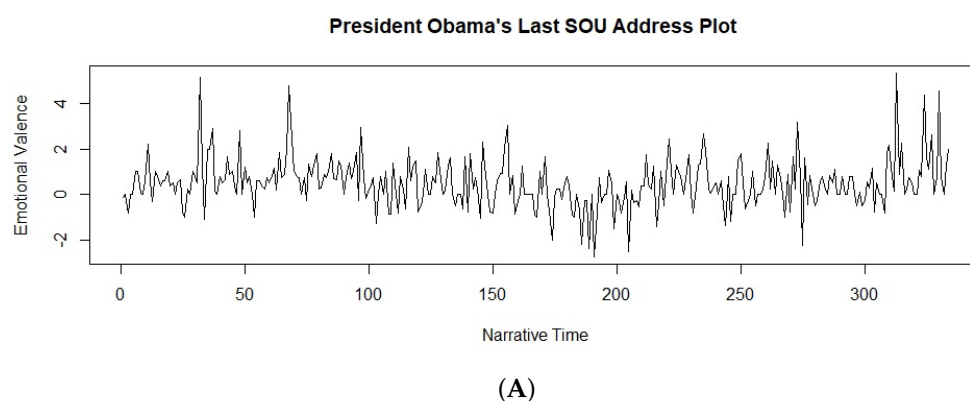


Another interesting feature of the four speeches is their “emotional valence”, or the pattern of sequential positive and negative emotions displayed as the speech unfolds through time. Plots of these patterns are shown in Figures 7 and 8. There is a distinct change in pattern in the emotional valence of President Trump’s two SOUs, as shown in Figure 7A,B. In the first, he commences on a positive emotional tone and is fairly upbeat in the first part of the speech, but then has multiple negative drops in the second half of the speech, before ending on a positive emotional note. In his second SOU, the pattern is roughly reversed, and there are more emotional negative points in the first half of the SOU, whereas the emotional volatility increases in the second half of the speech, with more frequent extreme highs and lows, and a predominantly positive tone at the end of the speech.

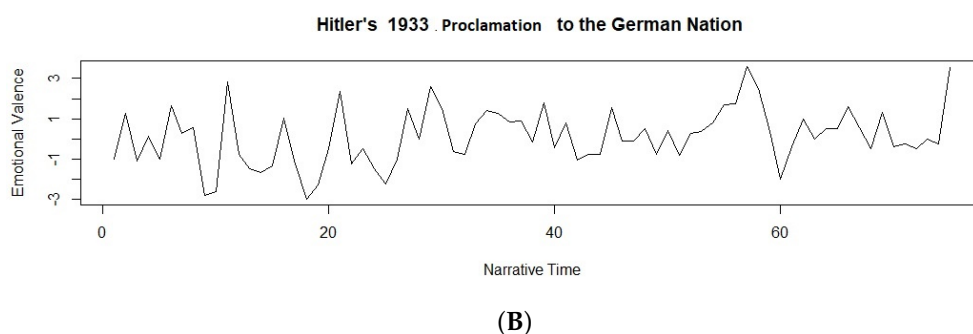
Figure 8A reveals that President Obama, in his last SOU, commences on a predominantly positive note, with some pronounced positive spikes, becomes more measured and negative in the middle of the speech, and ends on a predominantly positive note, with multiple positive peaks towards the end of his speech. Figure 8B shows that Hitler’s much shorter 1933 Proclamation is quite volatile in the first part of the speech, becomes more measured in the second half, with fewer extreme peaks and troughs, and finishes on a positive note.



**Figure 7.** The Emotional Valence of President Trumps two SOUs: (A) President Trump’s first SOU; and (B) President Trump’s second SOU.



**Figure 8.** *Cont.*



**Figure 8.** The Emotional Valence of President Obama’s last SOU and Hitler’s 1933 Berlin Proclamation: (A) President Obama’s last SOU; and (B) Hitler’s 1933 Berlin Proclamation.

### 3.2. Zipf Mandelbrot Analysis

Zipf [22] suggested that his “Theory of Relative Frequency” is a statistical law which falls within the laws of probability. Zipf’s law is an experimental law which is often applied to the study of the frequency of words in a corpus of natural language utterances. The law suggests that the frequency of any word is inversely proportional to its rank in the frequency table. In the case of the English language, the two most common words are “the” and “of”, and Zipf’s law states that “the” is twice as common as “of”.

Figure 9 shows plots of the application of Zipf’s law to the four speeches considered. The scales are in natural logarithms on both axes. A theoretical application of Zipf’s law would show a slope of negative one in the plots in Figure 9, running from top left to bottom right. All plots deviate from this concept, but the greatest deviation, from the theoretical concept, is in Hitler’s 1933 address, which is the most concave.

A flatter Zipf slope can indicate a more random signal, but it can also indicate a broader vocabulary that conveys a more precisely worded message. Zipf suggests that attempts to remove ambiguities should produce a flatter slope that favours the recipient. Mandelbrot [23] suggested that human languages have a slope of around 1. These political speeches are framed to favour the recipient. Hitler’s is the most extreme, but this is in translation. Obama’s is the closest to normal language, but is still some distance from it.

To further explore the degree of deviation in the context of these four speeches, we ran Ordinary Least Squares regressions of the log of rank regressed on the log of frequency. The results of these regressions are shown in Table 3.

The regression results in Table 3 reveal that all four regressions have F-Statistics that are highly significant, and Adjusted-R squares of 0.94, 0.94, 0.94, and 0.91, in the cases of President Trump’s two speeches, President Obama’s speech, and Hitler’s speech, respectively. The values of the slope coefficients, all of which are significant at the one per cent level, are Trump SOUA1 slope  $-0.67$ , Trump SOUA2 slope  $-0.71$ , Obama last SOUA slope  $-0.74$ , and Hitler 1933 slope  $-0.57$ .

These results suggest that all four political speeches are framed to favour the recipient. Hitler’s is the most extreme, but this is in translation. Obama’s is the closest to normal language but is still some distance from it. The most simplified and audience targeted is Hitler’s 1933 speech. Trump and Obama are close together, with Trump’s SOUAs showing slightly greater audience targeting.

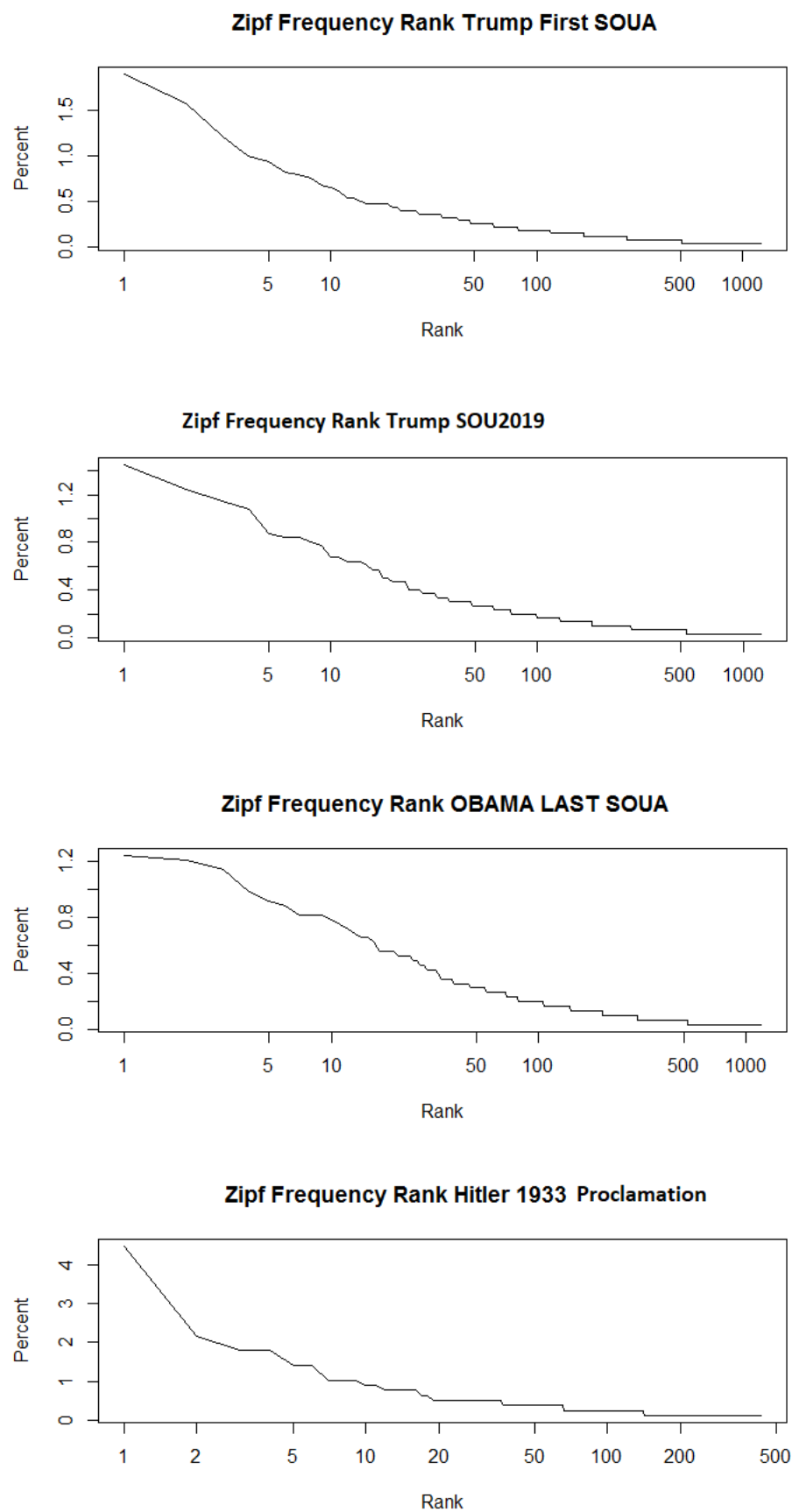


Figure 9. Zipf Plots.

**Table 3.** OLS, using Observations 1–1233 Dependent variable: l\_freqT1.

Coefficient		Std. Error	t-Ratio	p-Value
const	4.60818	0.0301961	152.6	0.0000
l_RankT1	−0.674643	0.00487040	−138.5	0.0000
Mean dependent var	0.478800	S.D. dependent var		0.687286
Sum squared resid	35.08483	S.E. of regression		0.168823
R <sup>2</sup>	0.939712	Adjusted R <sup>2</sup>		0.939663
F(1, 1231)	19187.55	P-value(F)		0.000000
Log-likelihood	444.8414	Akaike criterion		−885.6829
Schwarz criterion	−875.4484	Hannan–Quinn		−881.8328
OLS, using Observations 1–1227 Dependent variable: l_freqT2				
Coefficient		Std. Error	t-Ratio	p-value
const	4.83631	0.0306320	157.9	0.0000
l_RankT2	−0.706661	0.00494454	−142.9	0.0000
Mean dependent var	0.514392	S.D. dependent var		0.718456
Sum squared resid	35.80636	S.E. of regression		0.170967
R <sup>2</sup>	0.943419	Adjusted R <sup>2</sup>		0.943373
F(1, 1225)	20425.42	P-value(F)		0.000000
Log-likelihood	427.1954	Akaike criterion		−850.3908
Schwarz criterion	−840.1661	Hannan–Quinn		−846.5434
OLS, using Observations 1–433 Dependent variable: l_freqH1				
Coefficient		Std. Error	t- Ratio	p-Value
const	3.21372	0.0434618	73.94	0.0000
l_RankH	−0.565233	0.00840317	−67.26	0.0000
Mean dependent var	0.342408	S.D. dependent var		0.575778
Sum squared resid	12.45621	S.E. of regression		0.170002
R <sup>2</sup>	0.913026	Adjusted R <sup>2</sup>		0.912824
F(1, 431)	4524.478	P-value(F)		1.1e−230
Log-likelihood	153.8538	Akaike criterion		−303.7076
Schwarz criterion	−295.5661	Hannan–Quinn		−300.4936
OLS, using Observations 1–1189 Dependent variable: l_freqO				
Coefficient		Std. Error	t-Ratio	p-value
const	5.05132	0.0326043	154.9	0.0000
l_RankO	−0.740851	0.00528936	−140.1	0.0000
Mean dependent var	0.543524	S.D. dependent var		0.753179
Sum squared resid	38.44997	S.E. of regression		0.179979
R <sup>2</sup>	0.942946	Adjusted R <sup>2</sup>		0.942898
F(1, 1187)	19618.01	P-value(F)		0.000000
Log-likelihood	352.9148	Akaike criterion		−701.8296
Schwarz criterion	−691.6679	Hannan–Quinn		−698.0000

#### 4. Conclusions

In this paper, we have analysed President Trump’s two SOUs and contrasted the content with those of the last SOU of President Obama and that of Hitler’s 1933 Berlin Proclamation. All four are political speeches, and share a great deal of commonality. The sentiment analysis showed that they emphasize the nation, America and American, in the case of the two US Presidents, and Nation and German, in the case of Hitler. The word “will” features prominently in all four speeches, and relates to the respective political agendas of the speakers. The emotional tenor of the speeches of the two US Presidents is more positive than adopted by Hitler in his 1933 Berlin Proclamation. All speakers chose to end their speeches on a positive emotional note, and all four speeches contain Nationalistic elements.

The analysis also includes an application of the Zipf and Mandelbrot laws. The fact that all four had a slope of less than negative one, which would be standard speech in this framework, indicates that all three speakers had targeted their audiences and simplified the language used in their speeches. Hitler’s use of language was the most distant from standard speech with a score of negative 0.57. This suggests his status as a skillful mob-orator is justified. Presidents Trump and Obama were less extreme

but still had slope coefficients with values around negative 0.7, suggesting that they also target their audiences carefully.

The limitation of the text-mining approach adopted in the analysis of the contents of these four speeches is that it does not feature a verification of the statements made, and cannot pick up nuances in meaning and context. However, the approach does provide a broad indication of the structure and emotional flavour of the content, subject to the limitations of the lexicon applied. The Zipf analysis highlights the degree to which speech patterns within the speeches deviate from normal language values.

**Author Contributions:** D.E.A. and M.M. conceptualization and methodology, D.E.A. software, investigation and formal analysis, data curation, writing—original draft preparation, M.M. resources, writing—review and editing.

**Funding:** This research received no external funding.

**Acknowledgments:** For financial support, the first author acknowledges the Australian Research Council, and the second author is most grateful to the Australian Research Council; Ministry of Science and Technology (MOST), Taiwan; and the Japan Society for the Promotion of Science. The authors wish to thank four reviewers for helpful comments and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Allen, D.E.; McAleer, M. Fake News and Indifference to Scientific Fact: President Trump's Confused Tweets on Global Warming, Climate Change and Weather. *Scientometrics* **2018**, *117*, 625–629. [CrossRef]
2. Allen, D.E.; McAleer, M. President Trump Tweets Supreme Leader Kim Jong-Un on Nuclear Weapons: A Comparison with Climate Change. *Sustainability* **2018**, *10*, 2310. [CrossRef]
3. Allen, D.E.; McAleer, M.; Reid, D.M. Fake News and Indifference to Truth: Dissecting Tweets and State of the Union Addresses by Presidents Obama and Trump. *Adv. Dec. Sci.* **2018**, *22*. [CrossRef]
4. Feinerer, I.; Hornik, K. Tm: Text Mining Package. R Package Version 0.7-6. 2018. Available online: <https://CRAN.R-project.org/package=tm> (accessed on 27 August 2019).
5. Jockers, M.L. Syuzhet: Extract Sentiment and Plot Arcs from Text. 2015. Available online: <https://github.com/mjockers/syuzhet> (accessed on 27 August 2019).
6. Fellows, I. Wordcloud. 2018. Available online: <https://CRAN.R-project.org/package=wordcloud> (accessed on 27 August 2019).
7. Pröllochs, N.; Fuerriegel, S.; Neumann, D. *Understanding Negations in Information Processing: Learning from Replicating Human Behaviour*; Working Paper; Information Systems Research, University of Freiburg: Freiburg im Breisgau, Germany, 2017. Available online: <https://ssrn.com/abstract=2954460> (accessed on 27 August 2019).
8. Allen, D.E.; McAleer, M.; Singh, A.K. Machine News and Volatility: The Dow Jones Industrial Average and the TRNA Real-Time High Frequency Sentiment Series. In *Handbook of High Frequency Trading*, Gregoriou, G.N., Ed.; Academic Press: Cambridge, MA, USA, 2015; Chapter 19.
9. Allen, D.E.; McAleer, M.; Singh, A.K. An Entropy-based Analysis of the Relationship Between the DOW JONES Index and the TRNA Sentiment series. *Appl. Econ.* **2017**, *49*, 677–692. [CrossRef]
10. Allen, D.E.; McAleer, M.; Singh, A.K. Daily Market News Sentiment and Stock Prices. *Appl. Econ.* **2018**, doi:10.1080/00036846.2018.1564115 [CrossRef]
11. Tetlock, P.C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *J. Financ.* **2007**, *62*, 1139–1167. [CrossRef]
12. Tetlock, P.C.; Macskassy, S.A.; Saar-Tsechansky, M. More than Words: Quantifying Language to Measure Firms' Fundamentals. *J. Financ.* **2008**, *63*, 1427–1467. [CrossRef]
13. Da, Z.H.I.; Engelberg, J.; Gao, P. In Search of Attention. *J. Financ.* **2011**, *66*, 1461–1499. [CrossRef]
14. Barber, B.M.; Odean, T. All that Glitters: The Effect of Attention and News on the Buying Behaviour of Individual and Institutional Investors. *Rev. Financ. Stud.* **2008**, *21*, 785–818. [CrossRef]
15. diBartolomeo, D.; Warrick, S. *Making Covariance Based Portfolio Risk Models Sensitive to the Rate at Which Markets React to New Information*; Knight, J., Satchell, S., Eds.; Linear Factor Models; Elsevier Finance: Amsterdam, The Netherlands, 2005.



16. Mitra, L.; Mitra, G.; diBartolomeo, D. Equity Portfolio Risk (Volatility) Estimation using Market Information and Sentiment. *Quant. Financ.* **2009**, *9*, 887–895. [[CrossRef](#)]
17. Dzielinski, M.; Rieger, M.O.; Talpsepp, T. Volatility Asymmetry, News, and Private Investors. In *Handbook of News Analytics in Finance*; Wiley: Hoboken, NJ, USA, 2011; pp. 255–270.
18. Cahan, R.; Jussa, J.; Luo, Y. *Breaking News: How to Use News Sentiment to Pick Stocks*; MacQuarie US Research Report: New York, NY, USA, 2009.
19. Hafez, P.; Xie, J. Factoring Sentiment Risk into Quant Models, RavenPack International S.L. *J. Investig.* **2012**, *25*. [[CrossRef](#)]
20. Kearney, C.; Lui, S. Textual Sentiment in Finance: A Survey of Methods and Models. *Int. Rev. Financ. Anal.* **2014**, *33*, 171–185. [[CrossRef](#)]
21. Loughran, T.; McDonald, B. Textual Analysis in Accounting and Finance: A Survey. *J. Account. Res.* **2016**, *54*, 1187–1230. [[CrossRef](#)]
22. Zipf, G.K. *Selected Studies of the Principle of Relative Frequency in Language*; Harvard University Press: Cambridge, UK, 1932.
23. Mandelbrot, B. Information Theory and Psycholinguistics. In *Chapter in Scientific Psychology*; Wolman, B.B., Ed.; Basic Books: New York, NY, USA, 1965.
24. Shannon, C. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *379–423*, 623–656. [[CrossRef](#)]
25. Ficcadenti, V.; Cerqueti, R.; Ausloos, M. A Joint Text Mining-rank Size Investigation of the Rhetoric Structures of the US Presidents' Speeches. *Expert Syst. Appl.* **2019**, *123*, 127–142 [[CrossRef](#)]
26. Shklovsky, V. Art as Technique. In *Russian Formalist Criticism*; Lemon, L.T., Reis, M., Eds.; University of Nebraska Press: Lincoln, NE, USA, 1965.
27. Propp, V. *Morphology of the Folk Tale*; English Trans; First Published in Moscow in 1928; University of Texas Press: Laurence Scott, TX, USA, 1968.
28. Smith, A.D. Gastronomy or Geology? The role of Nationalism in the Reconstruction of Nations. *N. Natl.* **1994**, *1*, 3–23. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).