Sentiment Analysis on Public Statements during American Election Campaigns

Frederikke Pape Kofod (FPK)

School of Communication and Cognition, University of Aarhus

Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Kathrine Korsgaard Bredahl (KKB)

School of Communication and Cognition, University of Aarhus

Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Abstract (FPK & KKB)[1]

When a presidential candidate communicates, it is highly important to choose the words wisely to successfully convey the intended message. Depending on various variables, including social and economic events, the overall sentiment of these statements can vary. This research paper investigates, through sentiment analysis with a lexicon based approach, how the overall sentiment score differs in the two American political parties, the Democrats and the Republicans, during a campaign period prior to an election. This is examined by investigating the mean sentiment scores for each party throughout the period 1996 to 2016. Moreover, an interaction effect between the government in power and political party, seen in relation to sentiment score, is studied. There was found a significant difference between the sentiment scores of the two parties examined, and additionally the interaction effect between the government in power and party, was also found to be significantly different from the null-hypothesis. Factors like the universal positivity bias in language and negativity effects are

[1] All parts of the assignment were produced, composed and edited in agreement. The authors are represented alphabetically and by initials, independent of the degree of workload

included in the study, and lastly, confounding factors are discussed along with limitations of the sentiment analysis. Concurrently, different approaches and alternations are being suggested for future studies.

Keywords: Sentiment Analysis, Political Campaigns, LabMT, Positivity bias in language, Negativity effect

Introduction (FPK & KKB)

When navigating within the field of language, numerous people have tried to define the "function" of language. Several crucial questions have been raised, striving to investigate the relationship between language and thought - and how this affect cognition and our comprehension of the world. In a paper by O'Connell and Kowal (2003) it is outlined that psycholinguistics often have their primary focus on monologue. This disposition claims that monologue is solely about the person in whom the cognitive processes take place, making it fundamentally asocial. However, attention has made a move towards the importance of understanding the underlying dynamics in dialogue. Clark (1996) considers language as being collaborative, rather than focusing on its functional use. The contrasts in the theories are crucial for the conceptual comprehension of the "false dialogue situation" politicians navigate in, as these offer an alternative point of view on the basic mechanisms of linguistic interaction.

Political Language

Language, and the communicational aspect herein, is essential for the life we know. It is diverse, displaying itself purely and uncontrolled through spontaneous outburst, or intentional and meticulous as a tool. The latter is often experienced in the political arena, where politicians try to modulate their words to support their agenda most efficiently. These situations, albeit not considered as natural dialogues, still depend on theories relevant for the more traditional dialogue. A study conducted by Kuhlen and Brennan (2010) found results illustrating the importance of common ground between speaker and listener in a dialogue, and this seems to be applicable in political statements too, as the speaker might gain advantages by inferring the mental state of his or her listeners. Common ground and back-channeling appear to be beneficial for the speaker when trying to convey a message. Thus, to assure that a speaker’s words are suitable for a specific purpose, these are modulated by taking the mental state of the listener into consideration.

Political statements often have a lot at stake, but especially public statements during a presidential election campaign period are precarious. Besides the initial differentiation of opinion in politics, other factors might also affect the outcome of an election. Klein (1991) found that the personal dimension plays an important role in the battle for votes. The study found that negative traits have a greater impact than positive when evaluating a candidate running for president. This paper will investigate another underlying factor occurring during a campaign period, namely the language and the emotional value associated with it, when candidates express themselves and outline their beliefs. Today these expressions occur in various places during an election campaign. Communication on social media has become more and more common, also for political purposes, creating a huge amount of data to examine and interpret. Schoen et al. (2013) reported that this data has become an interest of many papers during recent years, because it generates new opportunities for explorations, for example within the field of predictions. Their study discusses electoral results as an research area of great interest, and mentions that despite sentiment analysis being a complicated and imperfect method so far, people still use it to predict electoral results.

Sentiment analysis

The goal of a sentiment analysis is to infer the psychological state of the author that has conducted the statement. As a field within Natural Language Processing, better known as NLP, it is used to experimentally understand the emotional content of a text. The theory is based upon the premise that words express sentiments, therefore by looking at the individual words in a text, an overall sentiment can be deducted. To infer the sentiment of a word, a sentiment dictionary can be used, containing a list of words with sentiment ratings. The approach is used in this paper, as it is computationally efficient. Some of the benefits of the method are that it uses corpus agnostics, which does not depend on training, and thus makes it a transparent procedure. (Dodds et al., 2015).

As mentioned earlier, the growing use of social media has led to an increased interest in studying the meaning of the utterances occurring on these platforms, as they are easily accessible and widely used by millions of people every day. A study conducted by Tumasjan, Sprenger, Sandner, and Welpe (2010) displayed an example of the use of sentiment analysis in tweets, as they investigated if Twitter is used as a forum for debates about political topics, and if the sentiment would resemble the real world's. The research found that Twitter is used as a platform to discuss political campaigns during the campaign period, and therefore the tweets might be relevant for the real world election status, grounding in the sentiment of the tweets.

However the paper by Tumasjan et al. (2010) has been criticized by Jungherr et al. (2011) for claiming that is possible, and even an easy process to predict an election, which is referred to in a paper of Gayo-Avello (2012). Gayo-Avello points out that predicting elections based on tweets has become a popular trend in recent years, because of the steep increase in the usage of social media. Additionally, he claims that when analyzing tweets using sentiment analysis, important factors like demographics and trustworthiness in the tweets are not taking into consideration, and thus makes the conclusions vague. However, this paper's investigation differs from the Twitter analyses because it focuses on longer utterances directly from the president candidates during the campaign period, rather than tweets anyone could have written, like scraping by specific hashtags. As Gayo-Avello mentions, the method of sentiment analysis is not capturing humor or underlying notes, especially not in tweets, as they are short and often full of nonverbal points. To try to compensate for some of the flaws in Twitter analysis, this paper reduces the area of interest to only analyzing longer statements from the candidates. This is also done based on the assumption that longer statements serve as better examples of instances containing elements from dialogue, like common ground, as the speaker is dependent on the immediate comprehension from the audience. Consequently, if a message is too vague, the delivered speech might not result in the wanted effect of a mutual connection.

LabMT

When doing sentiment analysis using a dictionary, a scale is needed to measure the weight of each word and whether they tend to be positive, negative or neutral. LabMT, the abbreviation for Mechanical Turk, is a dictionary tool serving this purpose. The dictionary scores happiness of a corpus based on the 5000 most frequently appearing English words on Twitter, New York Times, Google Books and music lyrics, creating a list of more than 10,000 words in total, that are then scored on their sentiment by Amazon's Mechanical Turk. The LabMT contains an "average happiness score", which is the polarity of the words (Reagan, 2017). Furthermore, it contains a happiness rank given the average happiness score, and the standard deviation for each word is to be found as well. Lastly, scores from the four sources is presented if possible. This paper uses the average happiness score, and as the LabMT operate on a “bag of words” approach, each word in the material is treated as an independent entity where context does not have any influence. As the words in this dictionary are frequent words, the material does not need to be tokenized or stemmed as other dictionaries require (Dodds, Harris, Kloumann, Bliss, & Danforth, 2011).

Negativity Effects in a positively bias language

Garcia, Garas, and Schweitzer (2012) suggest that language in general tend to lean towards a positive sentiment, perhaps to ease the process of creating social relations. Moreover, they claim that the usage frequency of words associated with positive emotions is higher than for words with negative associations. The idea of a general linguistic positivity bias is also supported by Dodds et al. (2015), who claim it to be a universal rule, including, but not limited to, the English language. Their research revealed that the English language appears to have a higher sentiment overall, compared to e.g. the Chinese and Russian. These results are crucial in further research reliant on sentiment classification, as former hypotheses have been based on the assumption that language is neutral. The contradicting discoveries advocated by Dodds et al. (2015) and Garcia et al. (2012) makes the interpretation of sentiment analysis more precise, as the universal positivity bias can be addressed and taken into consideration when explaining future research outcome. This result is also supported by Kloumann, Danforth, Harris, Bliss, and Dodds (2012) who, in addition to a universal positivity bias, also suggest that negative words are more diverse and influential, compared to positive words, substantiating a study by Garcia et al. (2012) examining how the degree of information in a word increases consistently in relation to the decrease in valance. Thus, a negative word contains more concrete meaning, than positive words. The variance in the impact of words with different sentiment is described by a theory defining the cornerstones of the phenomenon and explaining the manifestation of the effect.

Negativity Effect

The theory of negativity bias, sometimes also referred to as the negativity effect, describes how negatively charged pieces of information have a greater impact than equally positively charged pieces of information (Meffert, Chung, Joiner, Waks, & Garst, 2006). Additionally, this effect contributes to facilitate a change of opinion or enabling persuasion towards alternative concepts, when the given information is negative rather than positive, according to Allen and Burrell (2002). Negativity biases occur in several different occasions and are widely found in various aspects of everyday life. Several studies have been conducted, investigating the attributes of the negativity bias, focusing on its repercussions in different tasks involving information processing (Meffert et al., 2006). As an example, Kahneman and Tversky (1979) were interested in the phenomenon seen from an economic point of view, and defined the mechanisms of the negativity bias within economics as *loss aversion*, which, in summary, describes how losses have a greater impact than gains. The negativity effect also exemplifies itself in the Hindu caste system, where higher order caste members are more prone to "negativity contagion" than lower castes are to be by affected positively (Rozin & Royzman, 2001).

In addition to this, studies investigating the negativity bias have been able to observe it within numerous fields, relevantly in the department of political behavior too, as Meffert et al. (2006) mention.

The findings seem to suggest, that despite language in general tend to move towards a more positive sentiment, words with negative sentiment have a greater impact than positive ones. Thus, even though sentences typically are positive rather than neutral, sentences with a sentiment score lower than the average, seem to be more influential. This discovery might be central for candidates during campaign periods, as difference in sentiment could affect the outcome.

Lau (1985) conducted a study exploring the negativity effect on character perception in politics, and found that two main distinctions are essential in the explanation of how a change in the political environment might modify the way a politician is evaluated by the public. Lau suggests that the primary discrimination within the subject of the negativity effect should be made between “figure-ground negativity” and “cost orientation negativity”. He describes the figure-ground negativity as a perceptual approach emerging from the general positivity bias, resulting in negative information being perceived more noticeable, compared to the otherwise positive surroundings due to its less frequent appearance. Cost orientation negativity is outlined by a more motivational approach, concerned with avoidance of losses rather than approaching gains.

The defining characteristics indicates that cost orientation negativity, to a larger degree tends to be used intentionally during political campaigns, and several research papers have been interested in the dynamics of negativity bias within this area. As a natural extension from these results, it is of interest how presidential nominees are affected by this.

A study by Pinkleton, Um, and Austin (2002) investigated the effects of negative advertising in politics. Among other results, they found that the public attitude towards both what they describe as the “sponsor” and the “target” are evaluated more negatively in campaigns including negative advertising, which might indicate that a negativity bias is present, and how it possibly affects the public’s perception of candidates. Allen and Burrell (2002) also found evidence supporting this claim, and additionally suggested that politicians already utilize this, i.e. during “smudging” campaigns, where negatively biased stories are aimed to slander the reputation of an opponent. Ansolabehere and Iyengar’s work (1995) is mentioned in the paper by Allen and Burrell (2002) where an interesting point is made, finding that the inclination to vote decreased by almost 5 percentage points, when a potential voter saw a negative advertisement rather than a positive one, thus, indicating the presence of the negativity effect. Additionally, they refer to another study by Johnson-Cartee and Copeland (1991) that aligns with this finding, as they suggest that voter alienation inevitably is the most likely consequence following negative advertising during a campaign period. However, they hypothesize that a reverse effect of this is possible, causing voters to vote against a candidate that is associated with a possible negative outcome to prevent the advertised negative event from happening (Allen & Burrell, 2002).

Different wings, different politics – different sentiment?

The studies provide evidence for a universal positivity bias in the overall language sentiment, which creates the circumstances from which a contradicting negativity effect can emerge.

The papers mentioned earlier proposed that sentiment is affecting voters, and that studies have found that differences in sentiment do have an impact. These findings lead to other research questions, interested particularly with whether there is a connection between different political parties and their overall sentiment scores. The investigations of this paper are based on the fundamental political differentiation on norms, values, and beliefs between the Democratic party and the Republican party, aiming to examine whether these differences will affect the mean sentiment score for the two parties, respectively. The study also examines the development of the parties' mean sentiment scores during a campaign period leading up to a presidential election. The presidential elections of interest have been limited to the period from 1996 to 2016.

On these grounds, this study will investigate the following hypotheses:

1) We hypothesize that there is a significant difference in the mean sentiment scores of public statements during campaign periods and that the mean sentiment score will be higher for Democrats compared to Republicans.
2) We hypothesize that there is a significant impact on the average sentiment score in relation to the interaction effect between government in power and membership of the political parties in interest.

Materials and Methods (FPK & KKB)

The political statements were collected through several sources, mainly the web-database *The American Presidency Project*. These statements are all longer utterances, e.g. speeches, remarks and addresses. A corpus of 150 statements from the Presidential nominees for the United States Elections between 1996 and 2016 was obtained. In every election a minimum of six statements[2] from each presidential candidate were collected and these were distributed evenly throughout the campaign period. Because the United States currently do not have any rules restricting how soon a candidate can begin his or her campaign, some have a long duration time, and therefore this study has limited the span in which statements were

[2] With the exception of year 2000, where only 2 statements from Al Gore and 2 statements from Bush were accessible

withdrawn, to the 12 months leading up to election day, which always happens on the first Tuesday after November 1st (Moody, 2016). The materials for the first hypothesis consist of an equal number of statements from each party, resulting in 150 public statements distributed evenly between the Democrats and the Republicans. For the second hypothesis a specified search for public statements throughout the campaign period was conducted in order to detect the changes in the sentiment scores over time. The statements were manually derived and copied from the website and made into txt file on a computer. A Spyder (Spyder Developer Community, 1991) script was created in order to read the txt files and compute the sentiment score by using LabMT. To calculate the sentiment score of a statement an average happiness score is found by adding the scores in the "happiness_average" column of all the words and then divide this by the number of words. Then, for each rated word in the material, the script takes the word's score in the "happiness_average" column and subtracts the average score from it, ending with the sentiment score of the word. This is computed for all rated words in the dictionary. When all words possible have been rated these are put together and divided by the number of words in the material, rated or not - and this is the sentiment score of the material. The script composed all txt files into a single csv file, which was then used for analysis in RStudio (RStudio Team, 2015).

Analysis (FPK & KKB)

To test the hypotheses a sentiment score for each statement was calculated using the LabMT word list. Initially Spyder scripts were programmed for each hypothesis in order to create a unique csv file for each hypothesis. Both files were pre-processed in an Excel copy of the raw data. This pre-processing involved adding headers and separating the columns. Several statements from the corpus were transcriptions of public comments, and therefore all txt files were manually controlled for situation descriptions, like [laughter] which could affect the sentiment score profoundly, since this is the highest rated word in the LabMT word list. Additionally, minor pre-processing of the data regarding the first hypothesis was done in RStudio. A loop was made to create a new column containing the variable party by taking the first three letters of the txt file's name and adding this information to the column. This was done to make sure the categories within this were homogenized and thereby matched the sentiment scores. As all the assumptions for a t-test were met satisfyingly, an independent t-test was conducted, with sentiment score as the dependent variable, and political party as the independent variable. The standard deviation for the means of each party was also calculated. To get further statistical evidence the effect size for the test was computed. Lastly the visual illustrations of the results, the mean sentiment score for the Democrats and the Republicans, were created, using ggplot.

As the csv file for the second hypothesis was loaded into RStudio, it was possible to add a new column, in which the sitting government was presented appropriately to the sentiment scores. As the data did not violate the assumptions, a linear mixed effects analysis of the relationship between sentiment score and party was performed. The model was composed by the dependent variable sentiment score, and government and party as an interaction effect were the independent variables, the fixed effects. The dataset consisted of multiple data points from the same candidate, and so a random effect was added, to take the candidates' individual variance in sentiment score into account. The interaction effect between the two predictors was explored by conducting a pairwise post hoc test, in order to compare all different combinations of the independent variables. An example of these different combinations could be the sentiment score of the Democrats compared to the Republicans', when the sitting government is Democratic, as well a comparison of these scores, when the

Republicans are governing, and vice versa. The post hoc test was performed as Tukey's Honest Significant Difference, HSD. Again, visual illustrations of the results were produced by ggplot.

Results (FPK & KKB)

The independent t-test analysis of the data conducted to investigate hypothesis 1 found a significant difference in the average sentiment score between the Democratic party (*M* = 0.064, *SD* = 0.04), and the Republican party (*M* = 0.042, *SD* = 0.05), $t(143.2) = 2.8$, $p = .0059$, *SE* = 0.05 and $r = 0.22$. The results are illustrated in figure 1 and figure 2.

Figure 1: Mean sentiment score by party

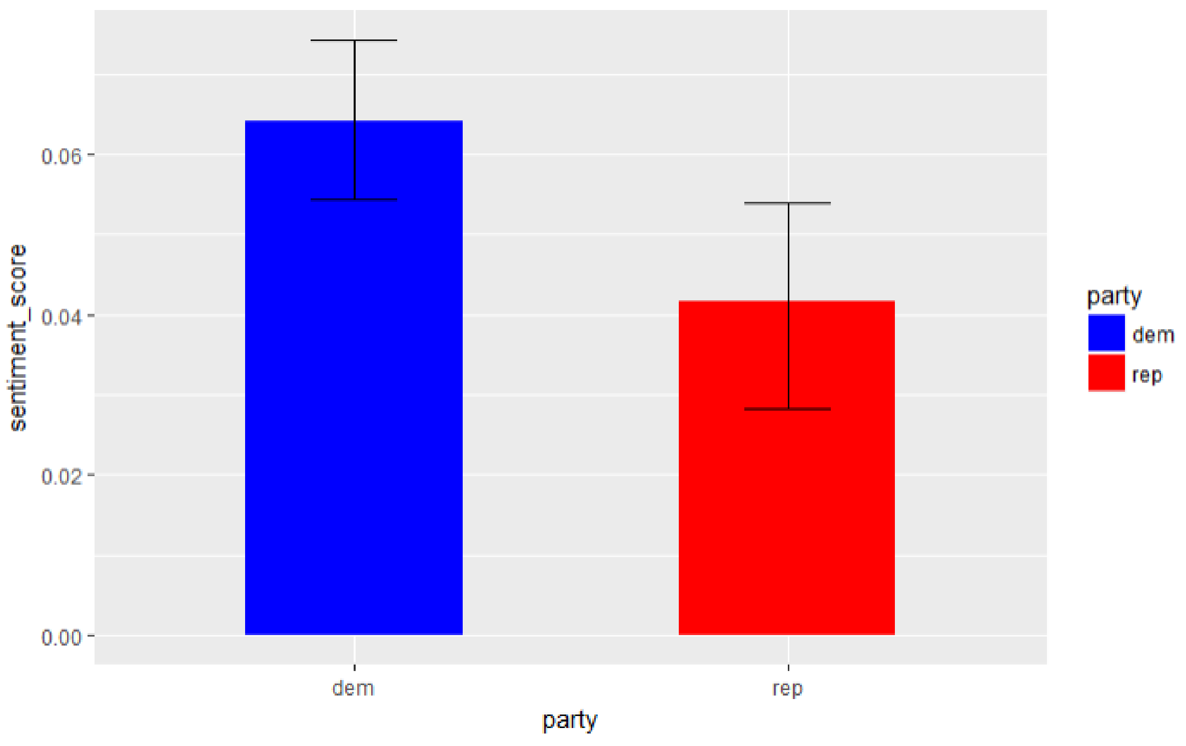
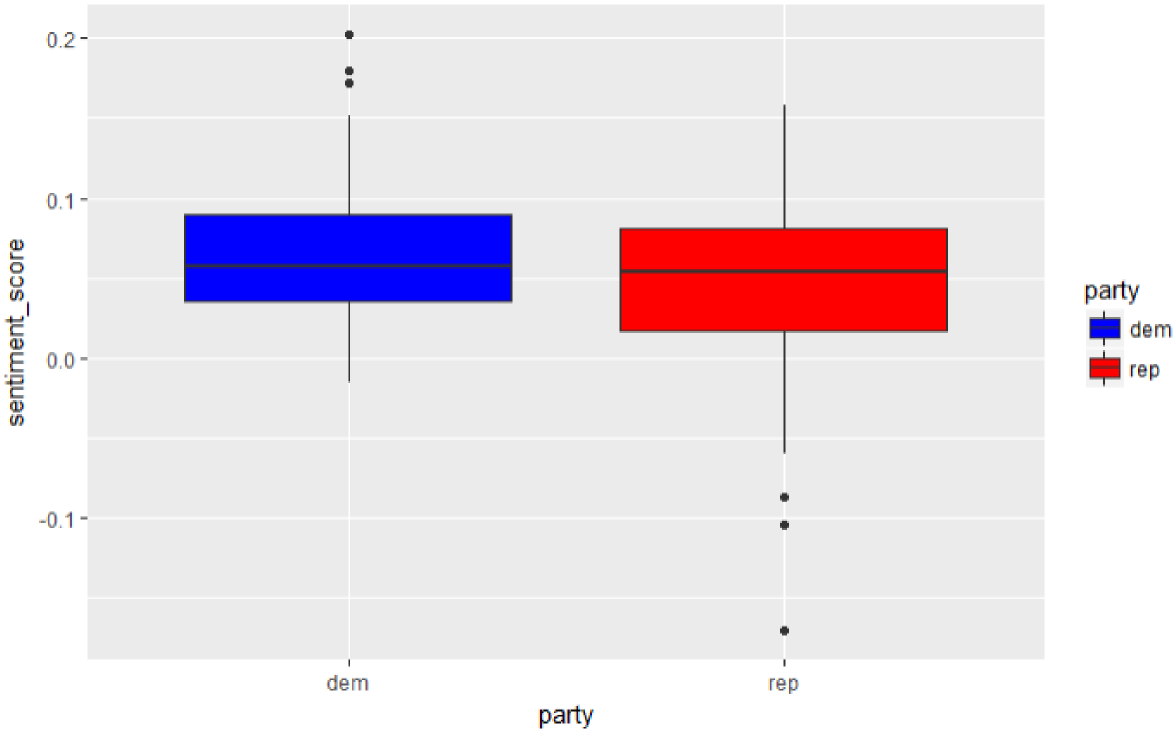


Figure 2: Mean sentiment score by party



To explore the second hypothesis a mixed effects model was used, with sentiment score as the dependent variable, government and party as fixed effects, and candidate as random effect.

sentiment score ~ government * party + (1|candidate)

The linear mixed effects model found that the interaction effect between the ruling government and the political parties significantly affected sentiment score, $\beta = 0.058$ ($SE = 0.023$), $t = 2.56$, $p = .013$. The interaction effect is illustrated in figure 3. Subsequently pairwise post-hoc tests were conducted using the “Tukey” method, but there was found no significant results. Figure 4 illustrates the mean sentiment score for the two parties, respectively to which party being the government in power.

Figure 3: Sentiment scores 1996-2016

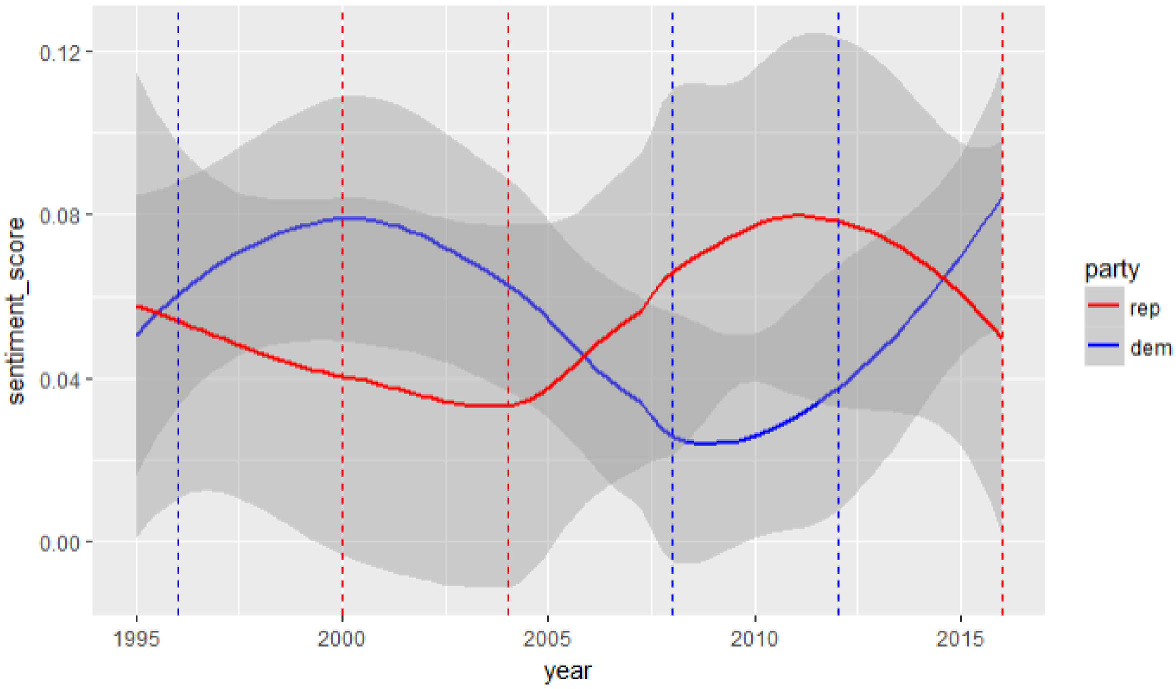
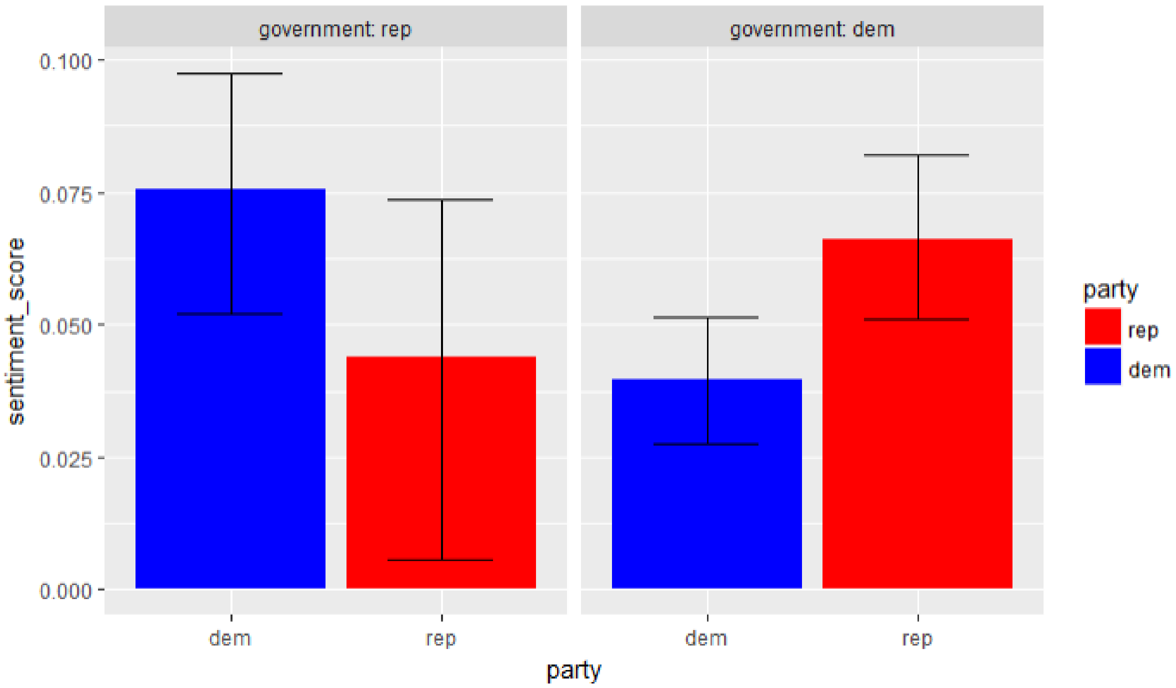


Figure 4: Sentiment score by sitting government



Discussion (FPK & KKB)

The investigation found a significant difference between the average sentiment score between the two parties, the Democrats and the Republicans, during the campaign periods between 1996 and 2016. This finding enables a rejection of the null-hypothesis, and thus has strong indications for the suggestion that the political differentiations between the examined parties will affect their average sentiment score. Additionally, it was found that the interaction effect between the government in power and which party a candidate is a member of, significantly affected the sentiment scores – albeit the exact place in which the means differed, was not specified. Despite this, the significant result of an interaction effect was found.

To address this issue a pairwise post hoc test was conducted, using the Tukey method. The test runs all pairwise contrasts of the fixed effects, government and party, and eliminates possible significant differences that might be due to overcorrection in the linear mixed model. It was found that none of the results were significant.

However, the visual inspections of the figures are in favor of the hypotheses. Figure 1 illustrates the mean sentiment score for each party, and demonstrates that the confidence intervals are not overlapping. The boxplot in figure 2 substantiates the results of the t-test, and the position of outliers support the initial thesis, that Democrats have a higher sentiment score than Republicans.

The interaction effect investigated in the second hypothesis is illustrated in figure 3. It appears as if both parties' sentiment scores are affected by which government is in power. The dashed vertical lines represent whether the elected president is a Democrat or a Republican. Thus, the figure indicates that when the government in power is Democratic, the

Democrats’ average sentiment score is lower than the Republicans’ average sentiment score, and vice versa. This is also illustrated in figure 4, although the confidence intervals are overlapping. Because the location of the effect was not significantly explained, supplementary variables might be influential. The results of the post hoc test might indicate that possible confounding factors also affect the sentiment score during a campaign period leading up to any given presidential election.

First off, the outcome of a presidential election is influenced by several different elements, whereof most of these are particularly difficult to fully grasp, making it hard to successfully transform these into measurable quantities. The outcome of an election will, hypothetically, always depend on the number of votes, but within the American electoral system, certain limitations and challenges can cause the outcome to suffer crucial consequences. Because the American voting system has two “layers”, both a direct Democracy in the popular vote, and as a representative Republic in the Electoral college, a presidential candidate who wins the popular vote is not guaranteed victory in the election. This is due to their voting system favoring winning the Electoral College, by getting a majority of the currently 538 votes (Huffington Post, 2012). The outcome of the latest presidential election was a demonstration of this, as Hillary Clinton won the popular vote by 55.3% to Donald Trump’s 46.09%, but lost the overall election, as Donald Trump won the Electoral College with 304 out of the 538, surpassing the line at 270. The complexity of the voting system might affect a candidate's public statements depending on whether they are giving a speech in a “safe state” or a “swing state”. Campaign speeches in swing states like Ohio or Florida often have more at stake, as the persuasion-wise aspect is more in focus. Therefore, the framework of the electoral system might have an influence on the strategic ideas behind the structure of campaigns, affecting the sentiment to an even larger degree (Revesz, 2016; Sachdev, Sehgal & Jasuja, 2017).

In addition to this, under ordinary circumstances the obvious, yet vital, variation in personal preferences within all people eligible to vote, has a major power. Besides the fact that people differ in their political point of view, their personal experiences and conception of the world is ambiguous, and thus, troublesome to make measurable and predictable. Even if it was possible to account for this, the manipulative grounds on which political campaigns often maneuver adds an additional layer of uncertainty, as the real intentions are challenging to discard from the pile of fake news, self-proclamatory PR, and falsified “public opinions” provided by opponents (Titcomb & Carson, 2017).

Allen and Burrell (2002) investigated how negative information as seen in smudging campaigns, affect the involved candidates, as well as how it in worst case scenarios threaten to undermine the basics pillars of Democracy, the voters. Pinkleton et al. (2002) also found, that the “sponsor” of negative advertising are perceived more negatively than the “target”, which might cause candidates to try to avoid making negative remarks. However, the political environment in the US is often associated with a high tolerance for critique, in which the opposing parties sometimes appear to be more concerned with bashing the rivalling party, rather than pursuing their own interests. The uncertainty of the incitement behind public statements complicates the attempt to clarify the strategic thoughts prior to them and thus, also the impact connected to the negativity effect.

Moreover, as stated in the introduction, Allen and Burrell (2002) mentions several studies, whereas the study by Johnson-Cartee and Copeland (1991) suggest that a reverse version of the presumed aftermath of the negativity bias in politics is plausible. If this is the case, the driving forces of the effect could be misleading for the comprehension of the outcome.

At the other end of the spectrum, supplementary uncontrollable circumstances might influence sentiment scores and election results. For instance, the fluctuations in both stocks as well as the housing market might have an unignorably impact on society. For starters, following such events the political scene is forced to act, and is expected to propose recovery plans for both economic aspects, but also fulfill the humanitarian responsibility. In the wake of a depression, both fiscal policy and economic policy are usually characterized by a revision of taxation, public budgets, labor markets, and interest rates, to do damage control. Following booms and busts, the government in power can choose to utilize different tools, i.e. either expansionary or contractive fiscal policy, which affects the sentiment of their public statements. This fundamental distinction generates an obvious dilemma, which often divide the waters (Amadeo, 2017).

Nevertheless, one’s political orientation does not seem to dictate a preference in fiscal policy. Examples of this could be that expansionary fiscal policy implemented by the Democratic Obama administration after the housing bubble in 2008, by a Republican government led by Bush in 2001 to handle the recession and boost the economy after the 9/11 terror attacks, and implemented in Democratic Franklin. D. Roosevelt’s “New Deal”, to end the Great Depression in the 1930’s. How a government or opposition proposes to deal with these kinds of events is most likely going to affect the sentiment of their public statements. Regardless of political orientation, they will have to align their message and word choice to the seriousness of the situation and address the core of the issue, making the use of negatively charged words inevitable. This example illustrates how unanticipated events in society can serve as a confounding factor when examining the general sentiment for political parties. The anticipated response is analogous, disregarding which party a candidate represents. This might result in a guideline for a mutual sentiment baseline, causing the statements to deviate from the sentiment usually occurring in unrestrained speech.

In addition to economic matters, the state of the country being at either war or peace, is ought to have an influence on the sentiment of politician’s utterances too. Statements given during a time of war will often include words with a very low sentiment score, like “terrorist”, “pain”, “death”. An example of this is seen in an excerpt from a statement by Bush in September 2004, which is included in the analysis material, saying: “This month in Beslan, we saw once again how the terrorists measure their success: in the death of the innocent and in the pain of grieving families”, which have a sentiment score of -0.247. In comparison, during a time of general domestic peace, these words will be less frequently used in public statements, and positive words used to retain status quo, occur more often. As in a statement by Obama, June 2008, saying: “That is why our heart swells with pride at the sight of our flag”, which has a positive sentiment score of 0.326. Thus, the difference between war and peace also serve as a confounding factor, influencing sentiment analysis.

Overall, the different factors, both sociocultural and society-wise need to be taken into consideration, as they will modulate the overall sentiments for both political parties. It is essential for the candidates to take a stand on the political topics, including unanticipated challenges. Thus, their “base sentiment” must conform to any unforeseen event, to keep up their personal trustworthiness, and avoid being perceived as inconsiderate or incompetent. This instinctive reaction adds insecurity to the analysis of sentiment in political statements.

Further investigations on the topic might include a variable containing a perspective on time, in order to capture variance in a more historical context. This would allow influential events like business cycles, natural disasters, and initiations of war to be compensated for and be taken into account in the examination of the outcome.

Although economic and social events could bias the results of this investigation, this might have been possible to prevent if more material would have been collected for the analyses. A larger data set would have decreased errors, and the issue of subjectivity too. As all materials has been manually selected, the process might have been biased by subjectivity. This has been tried to be accounted for by avoiding statements, e.g. at colleges or other special events that could impact the topic of the speeches and thereby the sentiment score. This could have been prevented if all the material had been available at the same place, and thus been collected without initial inspection. As the database used for this investigation had a large number of statements for most of the presidential candidates, others were unfortunately not in the database at all. To obtain a fairly balanced dataset with an equal number of statements per candidate, a compilation of several statements by internet search had to be done - and this manual search might have been biased.

A great limitation of the sentiment analysis is the bag of words principle as previously mentioned. When using this method in the LabMT dictionary for the analysis, several issues are not being addressed. One of the problems is mentioned by Gayo-Avello, who that political statements tend to contain irony, sarcasm and underlying messages (Gayo-Avello, 2012). This can influence the results of the analysis heavily, because the sentiment analysis approach cannot capture the usage of rhetorical devices. Since these are often used in public utterances, and especially in political ones including audience, it widens the margin of error. As the sentiment analysis also struggles with context sensitivity, the opinion mining can be incorrectly interpreted. Each word is an independent unit that has a separate polarity according to the sentiment analysis. When context does not have any influence, the negative constructed sentences will also be misinterpreted. This can change the average sentiment score of the statement or this party's beliefs, as the word "not" is not rated in the LabMT and will therefore contribute with a score of 0.0. Additionally, words with different meanings or connotations can affect the sentiment score as their context cannot be determined. An example of this could be the LabMT rated word "mean" which has multiple different interpretations - as an adjective for being evil and unfair, a noun for average, or a verb for thinking and considering. It is not described, which connotation the LabMT is rating. Nevertheless, the average happiness score for the word is 3.68, thus it is rated near the negative polarized words unfair, which is rated 3.34, and disgusting is rated as 2.58. This indicates that the LabMT give the words "mean" a fairly negative score, even when it is used in another context where it might be more positive. This illustrates the uncertainty about the sentiment analysis and the bag of words principle.

A way to prevent some of these precarious factors is described in a paper by Wilson, Wiebe, and Hoffmann (2005) which presents another dictionary approach to the sentiment analysis. This technique is called MPQA, short for the Multi-perspective Question Answering, and it automatically distinguishes the prior and contextual polarity. This means that the MPQA will take the context of the rated word into account, which will eliminate the negation problem in in the material examined. However, investigations have been conducted, where the MPQA did not perform very good, even though it includes the context. Islam and Zibran (2017) studied sentiment analysis with different dictionary approaches on Software Engineering material. They found that the dictionary SentiStrength performed the best in average accuracy, whereas the performance of MPQA was the worst of the four dictionaries investigated. The fact that this contextual approach is the least accurate dictionary is commented in their study, by the style of Software Engineering texts - these are often informal, where misspelled words or incorrect sentences can occur, however the MPQA is a universal purpose tool, and therefore it should be useful in the Software Engineering material. Even though Islam and Zibran conducted poor results for the context dictionary, this would still be an interesting method to include in future research purposes.

In a comparative study by Devika, Sunitha, and Ganesh (2016) different opinion mining methods are being investigated, examining both supervised and unsupervised approaches. Herein they describe the several levels of the sentiment analysis, as it can be used at word, sentence or document level. They examine the machine learning approach, where a trained algorithm can detect the sentiment of material; however, this approach requires a certain amount of training, before it can be not just useful, but also reliable. Additionally, the rule based method and the lexicon approach is being discussed, where the performance of the rule based is highly accurate in both “review level” and sentence level analysis. Nevertheless, the disadvantage is that the efficiency depends heavily on the rules created. The third and latter approach is the lexicon based, which is the approach used in this paper. The lexicon based method is unsupervised, and as discussed, it has several limitations, but is easy to approach and use. Devika et al. (2016) reports that the machine learning technique is the most accurate approach, and the field in which most work is being done. It would not have been possible to perform a machine learning algorithm for this investigation, since it requires enormous corpora and multiple training sessions on material similar to the one aimed to be examined. A different procedure might include looking at the sentiment of the candidates’ tweets to use the rule based approach or the MPQA method that includes the context of the words.

Conclusion (FPK & KKB)

This paper found the null-hypotheses to be rejected, as the t-test showed a significant difference in the average sentiment score between the Democrats and the Republicans within the investigated period, with the Democrats generally having higher sentiment scores than Republicans. Moreover the linear mixed effects model found that the interaction effect between the government in power and political party significantly affected sentiment score.

Acknowledgments

Thank you to Lasse Hansen for being a great tutor throughout this project, and to our teacher Kristian Tylén, for assisting us in the computational challenges.

References

Allen, M., & Burrell, N. (2002). *The negativity effect in political advertising*. Sage Thousand Oaks, CA.

Amadeo, K. (2017, August 26). Expansionary Fiscal Policy: Definition, Examples. *The Balance*. Retrieved December 13, 2017, from https://www.thebalance.com/expansionary-fiscal-policy-purpose-examples-how-it-works-3305792

Clark, H. H. (1996). *Using language*. Cambridge university press.

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44–49.

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., … Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, *112*(8), 2389–2394. https://doi.org/10.1073/pnas.1411678112

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). LabMT dictionary. Retrieved from http://journals.plos.org/plosone/article/file?id=info%3Adoi/10.1371/journal.pone.0026752.s001&type=supplementary

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS One*, *6*(12), e26752.

Fox News (2012, August 30). Transcript of Mitt Romney's speech at the RNC. *Fox News*. Retrieved December 5, 2017, from http://www.foxnews.com/politics/2012/08/30/transcript-mitt-romney-speech-at-rnc.html

Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative words. *EPJ Data Science*, *1*(1), 3.

Gayo-Avello, D. (2012). ‘I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper’ – A Balanced Survey on Election Prediction using Twitter Data. *ArXiv Preprint ArXiv:1204.6441*.

Huffington Post (2012, November 8). What is the electoral college? How it works and why it matters. *The Huffington Post*. Retrieved December 15, 2017, from https://www.huffingtonpost.com/2012/11/06/what-is-the-electoral-college_n_2078970.html

Islam, M. R., & Zibran, M. F. (2017). A Comparison of Dictionary Building Methods for Sentiment Analysis in Software Engineering Text. Retrieved December 18, 2017, from http://cs.uno.edu/~zibran/resources/MyPapers/DictionaryEvaluation.pdf

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk (2nd ed., Vol. 47, pp. 263–292). Econometrica.

Klein, J. G. (1991). Negativity effects in impression formation: A test in the political arena. *Personality and Social Psychology Bulletin*, *17*(4), 412–418.

Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PloS One*, *7*(1), e29484.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes, 47*(7), 567–587.

Lau, R. R. (1985). Two explanations for negativity effects in political behavior. *American Journal of Political Science*, Vol. 29, No. 1, 119–138.

Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, *56*(1), 27–51.

Moody, C. (2016, March 2). America's voting system is crazy - here's what you need to know. *CNN Politics.* Retrieved December 11, 2017, from http://edition.cnn.com/2016/01/31/politics/us-election-2016-explained/index.html

New York Times (2012, October 8). Transcript: Mitt Romney Remarks at Virginia Military Institute. *The New York Times*. Retrieved December 5, 2017, from http://www.nytimes.com/2012/10/09/us/politics/mitt-romney-remarks-at-virginia-military-institute.html

Nielsen, F. Å. (2015) LabMT. Retrieved from http://neuro.imm.dtu.dk/wiki/LabMT

O'Connell, D. C., & Kowal, S. (2003). Psycholinguistics: a half century of monologism. *The American Journal of Psychology*, *116*(2), 191–212.

Peters, G. & Woodley, J. T. (n.d.) The American Presidency Project. Retrieved from http://www.presidency.ucsb.edu/1960_election.php

Pinkleton, B. E., Um, N.-H., & Austin, E. W. (2002). An exploration of the effects of negative political advertising on political decision making. *Journal of Advertising*, *31*(1), 13–25.

Politico Staff (2012, August 10). Mitt Romeny's foreign policy speech (full text, video). *Politico*. Retrieved December 3, 2017, from https://www.politico.com/story/2012/10/mitt-romneys-foreign-policy-speech-full-text-video-082145?o=1

ProCon.org (n.d.) Retrieved from https://2008election.procon.org/view.resource.php?resourceID=001568#mccain

Reagan, A. (2017). Getting Started — labMTsimple 2.8.4 documentation. Retrieved 15 December 2017, from http://labmt-simple.readthedocs.io/getting-started.html#usage

Revesz, R. (2016, November 16). Five presidential nominees who won popular vote but lost the election. *The Independent*. Retrieved December 15, 2017, from www.independent.co.uk/news/world/americas/popular-vote-electoral-college-five-presidential-nominees-hillary-clinton-al-gore-a7420971.html

Romney, M. (2012, October 18). Mitt Romney's Speech at the Al Smith Dinner. *Real Clear Politics.* Retrieved December 6, 2017, from https://www.realclearpolitics.com/articles/2012/10/18/mitt_romneys_speech_at_the_al_smith_dinner_115842.html

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Sachdev, T., Sehgal, P., Jasuja, N. (2017, December 23). Electoral vote vs popular vote. *Diffen.* Retrieved December 17, 2017, from https://www.diffen.com/difference/Electoral_Vote_vs_Popular_Vote

Schoen, H., Gayo-Avello, D., Takis Metaxas, P., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*, *23*(5), 528–543.

The White House - President George Bush (n.d.) Retrieved from https://georgewbush-whitehouse.archives.gov/news/releases/2002/12/

Titcomb, J. & Carson, J. (2017, November 14). Fake news: what exactly is it – and can it really swing an election? *The Telegraph*. Retrieved December 12, 2017, from http://www.telegraph.co.uk/technology/0/fake-news-exactly-has-really-had-influence/

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Icwsm*, *10*(1), 178–185.

Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 347–354). Stroudsburg, PA, USA: Association for Computational Linguistics. https://doi.org/10.3115/1220575.1220619