Vol. 2, No.2, 2017

Journal of
language | works

Analysing Political Biases in Danish Newspapers Using Sentiment Analysis

Kenneth C. Enevoldsen & Lasse Hansen
Cognitive Science, Aarhus University
Kennethcenevoldsen@gmail.com; Lasseh0310@gmail.com

Abstract

Traditionally, the evaluation of political biases in Danish newspapers has been carried out through highly subjective methods. The conventional approach has been surveys asking samples of the population to place various newspapers on the political spectrum, coupled with analysing voting habits of the newspapers' readers (Hjarvard, 2007). This paper seeks to examine whether it is possible to use sentiment analysis to objectively assess political biases in Danish newspapers. By using the sentiment dictionary AFINN (Nielsen et al., 2011), the mean sentiment scores for 360 articles was calculated. The articles were published in the Danish newspapers Berlingske and Information and were all regarding the political parties Alternativet and Liberal Alliance. A significant interaction effect between the parties and newspapers was discovered. This effect was mainly driven by Information's coverage of the two parties. Moreover, Berlingske was found to publish a disproportionately greater number of articles concerning Liberal Alliance than Alternativet. Based on these findings, an integration of sentiment analysis into the evaluation of biases in news outlets is proposed. Furthermore, future studies are suggested to construct datasets for evaluation of AFINN on news and to utilize web-mining methods to gather greater amounts of data in order to analyse more parties and newspapers.

Introduction

The ideal of newspapers to be objective evolved in Denmark during the 20th century due to increased commercialization and professionalization, and though complete objectivity is virtually impossible to achieve, it has long been a prominent goal (Hjarvard, 2007). Hjarvard shows how political biases in Danish newspapers have been discovered using indirect measures, mostly through questionnaires asking about votes cast during the last election and which news outlet the voter preferred. Furthermore, questions asking people to place newspapers on a political scale from left to right wing have also been used (Hjarvard, 2007). These traditional measures reveal a tendency among people reading certain newspapers to vote for certain parties and that people tend to think of particular newspapers as representing particular political views. However, past research might be apt to produce slightly misleading results since few people keep or read all newspapers, and therefore rate them according to gut feelings or hearsay. Using the natural language processing method of *automated sentiment analysis* (Nielsen et al., 2011), this study aims to present a more objective analysis of potential political biases in newspapers. We tested, as a proof of concept, whether there is a more negative sentiment in articles concerning left wing parties published by what is generally considered a right wing paper, and vice versa. The main research question was thus whether it is possible to detect political biases in Danish newspapers using sentiment analysis. To test this research question, two of the most politically polarized papers, according to traditional methods, were chosen: *Berlingske* and *Information* (Hjarvard, 2007). In these two newspapers, articles concerning the chosen parties, *Alternativet* and *Liberal Alliance* were studied. The two parties were chosen since they were both relatively recently formed (2013 and 2008 respectively), have substantial media coverage, and represent different parts of the political spectrum, *Alternativet* being a left-wing party, and *Liberal Alliance* being a right-wing party.

Sentiment Analysis

Terminologically speaking, sentiment analysis is thought of as the "computational treatment of opinion, sentiment and subjectivity in text" (Pang & Lee, 2008, p. 1). Sentiment analysis enables the analyst to extract the emotionally laden parts of texts or corpora by analysing the lexis used in the text. Traditionally, sentiment analysis has been used on highly subjective texts with a strong valence such as tweets, reviews or personal blog posts. Sentiment analysis has been used to predict how well new movies will fare (Asur & Huberman 2010), how the stock market will behave (Bollen et al., 2011) and in revealing global positivity biases in natural human languages (Dodds et al., 2015).

Compared to newspapers, tweets are usually strongly emotionally laden, while newspapers are expected to live up to journalistic ideals of objectivity (Hjarvard, 2007). Therefore, it is uncertain how effective sentiment analysis proves to be at detecting political biases in newspapers.

The particular method of sentiment analysis utilized in this study uses a simple Python algorithm to query a human made corpus of words rated based on their emotional valence (hereafter a 'dictionary') to find mean sentiment scores (MSS) for a set of texts.

AFINN and the Dictionary Approach

The dictionary upon which the analysis is based, AFINN, is created by Finn Årup Nielsen and colleagues (Nielsen et al., 2011) and is one of a few dictionaries to include sentiment ratings of Danish words. The dictionary contains 3552 Danish words, all of them subjectively and singlehandedly rated by the principal creator on a scale from -5 (very negative) to 5 (very positive). Normally, a dictionary is constructed by multiple coders rating the same words to increase validity. This is clearly a limitation for AFINN and is further discussed in the discussion. It is worth noting that AFINN does not have ratings for 3552 unique words since many of them are simply inflections of the same word. AFINN works on the bag-of-words principle, which means that the text is represented as a bag wherein all the words are treated as independent entities. As such, co- and context are disregarded since it is assumed that the overall valence of the text is maintained. However, frequency of words is still kept. As a result of this, some preprocessing of the target materials is required. All articles need to be tokenized into individual words and all characters normalized to lowercase.[1] Some dictionaries require the normalization process to also include reducing words to their stem. This is not the case for AFINN, as it includes multiple inflections of the same word.

As an example of how AFINN works, an extract from the text corpus is shown below with the sentiment score of the individual words in parentheses. Positive numbers indicate positive words and, though none appear in this extract, negative numbers indicate negative words. Following this, a text with an overall positive sentiment, would receive a positive score.

> Anders Samuelsen er i gang med en opsigtsvækkende præstation. Sammen med sin nære allierede, Simon Emil Ammitzbøll, er han ved at lægge grunden til et solidt valgresultat og i fuld gang med at omdanne Liberal Alliance fra et protestparti til et parti, som vil søge indflydelse (2) og finde løsninger (1). Oven i hatten vil Anders Samuelsen - med en smule held (3) - kunne stryge forbi de Radikale på valgnatten. (Berlingske 2015-05-22)
>
> Anders Samuelsen is performing an astounding feat. With his close ally Simon Emil Ammitzbøll he is laying the foundation for a solid election result and at the same time converting Liberal Alliance from a protest party to a party, which seeks influence (2) and solutions (1). To top this off, Anders Samuelsen could – with a bit of luck (3) – pass Radikale Venstre on the election night. (translated from Danish by the authors)

This extract receives a mean sentiment score (MSS) of .091 (words not rated receives a score of 0[2]) which indicates that the overall sentiment of this extract is positive. The actual content of the extract is quite positive, since it is describing Liberal Alliance as being on the way to a good result in an election. As such, AFINN seems to properly capture the overall sentiment of the text. Looking at MSS for a single article can only tell whether the article is generally positive or negative, but by comparing scores across multiple articles it is possible to discern between levels of sentiment. As we will show in the Analysis, 0.091 constitutes a fairly strong positive sentiment score.

Dictionary based sentiment analyses such as AFINN have proven quite successful and powerful. When used to predict sentence polarity, *positive*, *negative*, or *neutral*, they have achieved accuracies ranging from 58.7% to 77% in certain studies, depending on the dataset used for evaluation (Koto & Adriani, 2015; Bravo-Marquez, Mendoza & Poblete, 2014), but other methods exist. It is worth noting that human annotators usually do not achieve greater inter-rater agreement on similar tasks than around 80% (Wilson, Wiebe & Hoffmann, 2005; Balahur et al., 2013). This means that only around 80% of texts can be confidently rated by human readers, and algorithms cannot be expected to surpass this barrier and, so to speak, detect sentiments that are not consistently detected by human readers.

This study is attempting to use an automated sentiment analysis to detect political biases in Danish newspapers. The underlying assumption being that if a newspaper has e.g. a negative bias towards a certain party, their articles about them will exhibit a negative sentiment, and vice versa.

In summary, the study sets out to test the following three hypotheses: 1) There are significant differences in sentiment for articles regarding *Alternativet* and *Liberal Alliance* in the newspapers *Berlingske* and *Information*, 2) articles concerning *Alternativet* exhibit a more positive sentiment in *Information* compared to *Liberal Alliance*, and 3) articles concerning *Liberal Alliance* exhibit a more positive sentiment in *Berlingske* compared to *Alternativet*.

Materials and Methods

Through the use of the Danish web-database, *Infomedia*[3], a corpus of 360 articles from the printed newspapers *Berlingske* and *Information* which regarded the political parties *Alternativet* and *Liberal Alliance* was manually compiled. An equal number of articles from each newspaper about each party was included. The four categories constitute the four conditions.

For each condition 30 articles were included in the time period from the 1st of January 2015 through the 31st of June 2015; 30 in the period from the 1st of July 2015 through the 31st of December 2015; and 30 more in the period from the 1st of January 2016 through the 30th of June 2016, for a total of 90 articles from each newspaper regarding each party. This was done to include comparable coverage of the two parties by the two papers under the assumption that 30 articles from each 6-month period would constitute the most relevant articles in each time period. The corpus only included articles wherein the exact name of the party appeared and excluded any articles wherein both parties were mentioned. Furthermore, articles shorter than 60 words were excluded, since these were often nothing but a reference to other articles or front page texts. Since one of the chosen parties' name, *Alternativet*, is also a Danish word (meaning 'the alternative') articles in which the word was not used to describe the party were manually excluded. The articles in each time period were sorted by *Infomedia*'s relevancy filter, which uses frequency of the search input, the search input's frequency in the overall database, the length of the article as well as other factors[4]. In each search, 30 articles which fit the criteria were chosen for each condition.

Analysis

To test the hypotheses, the mean sentiment score (MSS) for each article was calculated using the AFINN Danish dictionary. Using Python, all articles were tokenized into individual words and all characters normalized to lowercase. The MSS was calculated as opposed to the absolute sentiment score for each article to get a normalized value which would be comparable across articles with different lengths and numbers of rated words. This resulted in MSS ranging from -0.11 to 0.12 with a global mean of 0.0085. The reason for these seemingly low numbers is that the greater majority of words present in the articles are neutral or not rated, and as such have a sentiment score of 0. Note that the global mean is above zero even though AFINN contains a greater number of negatively rated words than positive (65.5% negative). This might indicate that newspapers, in general, tend to shy away from overtly negative language.

Based on the hypotheses, an interaction effect between parties and newspapers is expected. Using MSS as the outcome variable, a two-way factorial ANOVA was conducted with party and newspaper as predictor variables. This was additionally followed up by pairwise Bonferroni corrected post-hoc tests to further examine any potential political biases found. Both statistical tests were conducted using RStudio (RStudio Team, 2015).

Results

We found no main effect of newspaper on the MSS of articles, $\beta = 0.008$ ($SE = 0.003$), $t(356) = 1.632$, $p = .104$. There was also no main effect of party on the MSS of articles, $\beta = 0.003$ ($SE = 0.005$), $t(356) = 0.584$, $p = .560$. This means that there is no overall significant difference between the MSS of the articles from the two newspapers and no overall significant difference between the MSS of the two parties either. In other words, neither of the two papers consistently write more positively than the other paper; and neither of the two parties is consistently described more positively than the other.

However, supporting our predictions, we found a significant interaction effect between newspaper and party, $\beta = -0.16$ ($SE = 0.007$), $t(356) = -2.35$, $p = 0.019$. This means that there is a significant difference in the way the newspapers report on the two parties.

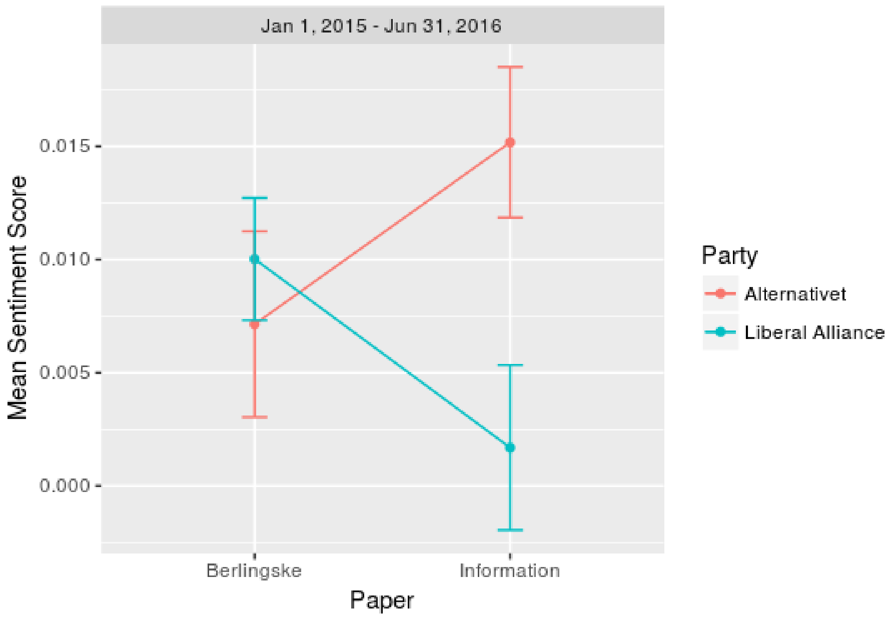

Figure 1: Mean Sentiment Score (MSS) by newspaper and party.

Pairwise Bonferroni corrected post-hoc tests revealed this interaction to be mainly driven by *Information*'s coverage of the two parties. The MSS of articles in *Information* regarding *Liberal Alliance* was 0.0017 (SD = 0.035), while articles regarding Alternativet was 0.0152 (SD = 0.032). This effect was significant, $p = .039$, d = 0.41, r = 0.2.

Figure 1 shows the MSS for each party as a function of which newspaper published the articles. A clear interaction effect is shown by the crossing lines, indicating a significant difference in the sentiment scores for *Alternativet* and *Liberal Alliance* in *Information*. The overlapping errorbars show that *Berlingske* displays no significant difference in the sentiment scores of articles concerning either party.

Discussion

The results suggest that there is a significant difference between the coverage of *Alternativet* and *Liberal Alliance* by the two newspapers, *Information* and *Berlingske*. This supports our first hypothesis that a party will be presented more positively by a newspaper with a political bias towards the party's wing. Pairwise post hoc tests reveal this difference to be mainly driven by *Information*'s coverage of the two parties. It is observed that *Information*'s articles generally have a higher MSS when concerning *Alternativet* compared to *Liberal Alliance*, which is in line with the second hypothesis. The third hypothesis cannot be confirmed, since there is no significant difference in MSS for articles in *Berlingske* regarding the two parties.

On the basis of these findings, it is suggested that an automated sentiment analysis can be used to obtain a more objective and quantitative measure of the biases inherent in newspapers. However, some methodological concerns need to addressed before a large-scale analysis can be performed. These include problems regarding co- and context, the dictionary of AFINN and the lack of a validity test. To paint a more accurate picture of political biases in newspapers, more than two parties should be included in the sentiment analysis.

Though *Berlingske* might not display a bias in their sentiment, another factor is worth investigating. As shown in table 2, *Berlingske* published 63% more articles on *Liberal Alliance* than *Alternativet* (781 and 479 respectively) in the time period from 2015-01-01 through 2016-06-30. For comparison, *Information* only published 11% more articles on *Liberal Alliance* than *Alternativet* (341 and 307 respectively). Earlier studies have suggested that a way to find biases in news, is to look at what facts the media choose to present and which they do not (Balahur & Steinberger 2009; Balahur et al. 2013). Simply not reporting on a topic or a party is thus a way in which a possible political bias can be expressed in newspapers. Furthermore, as mentioned in the introduction, Duan & Whinston (2008) points out that amount of reviews is a good indicator of movie sales. More generally speaking, if something is mentioned often, it is perceived as being more true and important than things that are mentioned less often (Tversky & Kahneman, 1973). If this is transferable to news outlets as well, it could indicate a political bias not only reflected in sentiment score, but also in representation.

Table 2: Total Number of Published Articles concerning the parties (2015-01-01 – 2016-06-30)

	Alternativet	Liberal Alliance	Total articles about the two parties
Berlingske	479	781	1260
Information	307	341	648

A few more points are worth bringing up. Firstly, *Liberal Alliance* received 57% more votes than *Alternativet* in the 2015 general election (7.5% and 4.8% of total votes respectively[5]). It is, therefore, debatable whether *Berlingske*'s disproportionally greater reporting of *Liberal Alliance* is simply a reflection of this difference in votes or a bias. This factor might also be what is driving *Information's* greater coverage of *Liberal Alliance*. Notice that a bias in sentiment might have been shown in *Berlingske* had other parties been examined. Further studies examining a greater number of parties are needed to fully assess both sentiment and representation biases in Danish newspapers.

Methodological Limitations

As previously mentioned, the bag-of-words approach to sentiment analysis disregards co- and context, which potentially causes some problems to arise. For instance, irony, sarcasm and who is speaking to whom is not taken into account. A consequence of this is that articles which are expressing a party's negative opinion towards a subject (be it a new bill, political opponent or otherwise) will attach a negative sentiment score to the party. This is a source of error, which cannot be easily remedied. A potential solution could be a machine learning algorithm supplied with a massive dataset of rated articles and texts snippets, or simply by excluding citations from the analysis. This will be further examined in the following sections. In this analysis we rely on the assumption that the statistical noise introduced by this error is equally distributed across conditions. This leads to a decrease of statistical power, but does not invalidate the results.

Still, some other methodological problems remain. As an example, an extract from an article in the text corpus is shown here:

> Jeg har ingen tillid (2) til, at organisationen stopper (-1) sig selv i forhold til at producere nye skandaler (-3), « siger Merete Riisager og fortsætter: »Kombinationen af et højt skattetryk og et dysfunktionelt skattesystem er hamrende farlig (-2), fordi borgerne mister (-2) tilliden til Skat, når der er eksempler på uretfærdigheder i systemet. (Berlingske 2016-01-16)
>
> I have no trust (2) that the organization will stop producing new scandals (-3)« Merete Riisager says and continues: »The combination of high taxes and a dysfunctional tax

> system is extremely dangerous (-2), because citizens lose (-2) their trust in Skat [the Danish tax system] when there are examples of injustices in the system. (translated from Danish by the authors)

This extract from a *Berlingske* article about *Liberal Alliance* illustrates some general flaws in sentiment scoring. First of all, the phrase "*I have no trust that…*" is rated as positive, even though the adverb '*no'* reverses the polarity of the sentence. Furthermore, and most importantly, the article is not directly critical of *Liberal Alliance* but quotes a member of the party stating an opinion on the tax system. This produces an MSS for this extract of -0.125, which is a very low score compared to the general MSS presented in the text corpus (see analysis). This is despite words such as '*dysfunctional'* and '*injustices*' not being rated. The automatic sentiment analysis will evaluate the extract as highly negative of *Liberal Alliance*, whereas a contextually sensitive reading will see it as neutral (i.e. not presenting any stance towards the party).

Furthermore, AFINN contains only 3552 rated words and therefore in some cases entire paragraphs are rated positive or negative based on a single or very few words. Another pertinent issue with AFINN is that the creator, Finn Årup Nielsen, single-handedly rated all of the words. Therefore, validity cannot be reviewed, which is a major source of error using AFINN. Furthermore, AFINN was created mainly for microblogging analysis, namely for tweets (no more than 140 characters). As a result of this, many obscene words and expressions which would never occur in news articles are also included in the word count while many words that are common in the language style of newspapers are missing. Furthermore, it is important to stress again that AFINN does not contain 3552 unique words, since many of the words simply are inflections of the same word. For instance, 5 different inflections of 'ødelægge' (to destroy) are included.

To remedy some of AFINN's most obvious flaws, one of the easiest steps would be to expand the dictionary size to better account for words apparent in newspapers. However, this will not correct problems inherent in the bag-of-words approach. Other sentiment analyses have included cotext into the scoring of words (Wilson et al., 2005) e.g. by making negations reverse the sentiment of the following word. Furthermore, one could have adjectives and adverbs which modify the sentiment valence of a word, such as 'very' and 'much', increase or decrease the sentiment score. In a more complex approach it might also be possible to decrease sources of error derived from context by excluding or treating quotations separately. However, it is unclear whether or not any of these interventions will significantly improve the result of the analysis.

Alternatives to AFINN

As previously mentioned, AFINN is one of a few sentiment dictionaries to include Danish words, but other dictionaries bypass this using tools such as *Google Translate*. One of the most prominent dictionaries utilizing this is the *NRC Emotional Lexicon*. NRC uses emotion categories such as *fear*, *anger* and *trust* to classify words. NRC has some problems which include mistranslation and cultural differences in word sentiment. For instance, the word "socialist" is associated with *fear*, *anger*, *sadness* and *disgust* in the NRC dictionary, which we assume would not be the case in a Danish cultural context. It is, therefore, unlikely that NRC would be applicable for this type of research.

As previously mentioned, an alternative way to approach the analysis could be by manually creating a training dataset for a machine learning algorithm. Coders could have read through and labelled e.g. 120 articles and used these as the basis for the scoring instead of AFINN. This labelling would entail marking e.g. each paragraph as being positive, negative or neutral. A simple algorithm can then be made which tries to predict the polarity of new sentences by looking at how frequently the words in it appear in each category. If the words appear more often in positively labelled sentences, it would classify the sentence as being positive and so forth. This method could quite easily be expanded by having the algorithm detect inverters such as ‘not’ and modifiers such as ‘very’. This method would essentially create a new dictionary which would be specifically tailored for news.

More complex machine learning algorithms such as deep learning could be trained to predict human ratings. This algorithm would ‘learn’ to predict human ratings by detecting patterns in text which has been classified as positive or negative. This more complex machine learning method would theoretically be able to take co- and context into consideration, although we would be unable to know how a certain sentiment score would have been produced, essentially making the rating of the text into a black box. Besides the black box problem, the deep learning approach would require immense amounts of rated data (in the order of at least 1000-10000 rated data points) before it becomes a viable alternative.

A major problem with these machine learning methods, is that they require a lot of work and a large dataset to be trained on before they reach reasonable performance. No sentiment labelled dataset of Danish news exists to our knowledge, and it was found beyond the scope of this study to construct one. For future research on this topic, it might be advantageous to construct such a dataset not only for algorithm training but also for validation of results. Contrary to machine learning methods, the dictionary approach works without any training and can be applied immediately.

Future Directions

As previously mentioned, AFINN was constructed for sentiment analysis of tweets, which are short, to the point and usually quite valent. As such, evaluation of AFINN’s scores has mainly been done on datasets of tweets and to a smaller extent movie reviews[6]. It might not necessarily fare as well at other types of text. It can be argued that a different approach might be necessary when dealing with news text since the target of the article is less obvious and different sentiments towards different entities might be expressed in the same article (Balahur & Steinberger, 2009; Balahur et al., 2013).

All articles used for the analysis were collected manually from *Infomedia* and all were found using simple search mechanics. Only a few cases were excluded manually e.g. cases in which the Danish word, 'alternativet', was not used to describe the party (see 'Materials and Methods'). These cases were few and could have been minimized using a better search engine to only include cases where 'alternativet' appeared with a capital A and not after a full stop. Therefore, this study could have increased statistical power by using web-mining to obtain a greater number of articles. This could have been done using *Infomedia*'s API, which was, unfortunately, inaccessible. Using this method, other parties as well as newspapers, could be examined for political biases toward the right or left wing.

Conclusion

This paper analysed articles from the Danish newspapers *Berlingske* and *Information* and found a significant difference in the way they portray the political parties *Alternativet* and *Liberal Alliance*. Using the sentiment dictionary AFINN, the mean sentiment score (MSS) for a total of 360 articles in the time period 2015-01-01 through 2016-05-31 was calculated. The results reveal *Information*'s articles mentioning *Alternativet* to be more positively worded than articles mentioning *Liberal Alliance*. The results also indicate *Berlingske* to be less politically biased in the sentiment of their articles, showing no significant difference in their portrayal of the two parties. However, *Berlingske* published 63% more articles on *Liberal Alliance* than *Alternativet* in the time period.

The study's main contribution is a novel approach to using sentiment analysis as a tool for obtaining more objective and quantitative measures of the political biases inherent in newspapers. This also opens the opportunity for the use of sentiment analysis in evaluation of objectivity in e.g. state-sponsored media outlets. There are some limitations in methodology in regards to the dictionary and bag-of-words approach, which warrant future research efforts. Especially, ways to incorporate co- and context into the sentiment rating should be examined. Furthermore, this study only analysed articles concerning two parties. More parties should be evaluated before the biases of newspapers can be affirmed. Future studies can be carried out to investigate the effectiveness of AFINN at scoring news by creating an evaluation dataset, and by using web-mining techniques to acquire a larger dataset upon which more parties and newspapers can be evaluated.

References

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology,* Vol. 1, 492–499. https://doi.org/10.1109/WI-IAT.2010.63

Balahur, A., & Steinberger, R. (2009). Rethinking Sentiment Analysis in the News: from Theory to Practice and back. *Proceeding of WOMSA, 2009.* Retrieved from http://www.scss.tcd.ie/Khurshid.Ahmad/Research/09_WOMSA-WS-Sevilla_Sentiment-Def_printed.pdf

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., van der Goot, E., Halkia, M., … Belyaeva, J. (2013). Sentiment Analysis in the News. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010),* 2216-2220. Retrieved from http://arxiv.org/abs/1309.6202

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, *69*, 86–99. https://doi.org/10.1016/j.knosys.2014.05.016

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

Duan, W., Gu, B., & Whinston, A. B. (2008). Do online reviews matter? — An empirical investigation of panel data. *Decision Support Systems*, *45*(4), 1007–1016. https://doi.org/10.1016/j.dss.2008.04.001

Hjarvard, S. (2007). Den politiske presse - En analyse af danske avisers politiske orientering. *Journalistica - Tidsskrift for forskning i journalistik*, *2*(5). Retrieved from http://ojs.statsbiblioteket.dk/index.php/journalistica/article/view/1808

Koto, F., & Adriani, M. (2015). A Comparative Study on Twitter Sentiment Analysis: Which Features are Good? In *Natural Language Processing and Information Systems*, 453–457. Springer, Cham. https://doi.org/10.1007/978-3-319-19581-0_46

Nielsen, F., Rowe, M., Stankovic, M., Dadzie, A., Hardey, M. (2011). A new ANEW: evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. Volume 718 in CEUR Workshop Proceedings,* 93-98.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, *11*(122–129).

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *LREc* Vol. 10, 1320-1326.

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135. https://doi.org/10.1561/1500000011

Rstudio Team (2015). Rstudio: Integrated Development for R. Rstudio, Inc., Boston, MA. URL: http://www.rstudio.com

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

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,* 347–354. https://doi.org/10.3115/1220575.1220619

Ye, Q., Law, R., Gu, B., & Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, *27*(2), 634–639. https://doi.org/10.1016/j.chb.2010.04.014

Noter

1 Contractions are all rated as unique words. This means, that even if part of the contraction is rated, it will not influence the rating of the contraction. E.g. if ”pistol” (gun) has a rating, it will not influence the rating of other words containing ”pistol” such as ”pistolhylster” (gun holster).

2 See Analysis for more information on calculation of MSS.

3 http://infomedia.dk/mediearkivet/

4 For more information on Infomedia’s filters: https://support-infomediadk.ez.statsbiblioteket.dk:12048/da/products/mediearkiv/soegeresultat/

5 Election result: http://www.dr.dk/nyheder/politik/valg2015/resultat

6 See http://neuro.imm.dtu.dk/wiki/AFINN#Evaluation for a complete overview.