Analysis of the online Swedish political discourse
Thesis report for Degree Project C in Computer Science

Pelle Ingvast
Abstract

Analysis of the online Swedish political discourse

Pelle Ingvast

In any written text bias is inherent, therefore it is important as a consumer of media to understand that there are underlying biases. Sweden have a multi party system in parliament meaning that bias in media is possible not only in favor of two ideologies but potentially eight. Automatically detecting bias or political leaning in news is seldom done for other languages than English.

This project attempts to automatically detect the political bias and leaning of Swedish newspapers. Two models are proposed and evaluated for two types of bias, visibility bias and sentiment or tonality bias, one model for each type of bias. Four of the biggest Swedish newspapers are evaluated using the aforementioned models. The results are then compared to an expert opinion on the political leaning of the newspaper. Visibility bias was conducted on an article level whereas sentiment or tonality bias was determined on a sentence level. For sentiment analysis the Swedish port of Vader was used. The English version of Vader has historically struggled with editorial pieces this was true for this project as well, despite breaking up the pieces into more digestible sentence blocks.

The results found that the visibility bias of each party reflects its size. In other words parties with more mandates gets more coverage. The predictive capabilities of the visibility bias model is only applicable in the case of significant outliers. For example Sverigedemokraterna scored higher than the average visibility bias in 3 out of 4 newspapers. Furthermore after the election Sverigedemokraterna had the largest percentile increase in votes. However Miljöpartiet scored the second highest of the smaller parties in visibility bias across 3 out of 4 newspapers, although in all cases below the average visibility bias. Yet Miljöpartiet was the only of the smaller parties to lose voters.

As for sentiment analysis, it indicates that all newspapers tested are on average neutrally or positively inclined towards all political parties. However since the range of the results was smaller than expected it is hard to definitively draw any conclusions. Suggestion for future improvements on the model as well as suggestions for other approaches are provided.
1 Introduction

In the last decade newspapers have shifted to the web. This in turn have made large quantities of data available for analysis, therefore a change has occurred where the once social-science related issue of bias has been opened up to the application of big data analysis [1].

This project attempts to automatically quantify the bias in Swedish online newspapers to determine if there exists any inherent bias within them. Furthermore, this project evaluates the models and methods used to determine the bias. The evaluation of the models and methods is done by cross-referencing the results with expert opinions on the bias of the newspaper.

However, before going into too much detail to understand this one must first fully grasp the meaning of bias. Ironically bias might be most easily understood by its opposite, objectivity, essentially objectivity is the complete lack of bias. Whenever subjective elements are introduced bias becomes apparent. For example if I were to write an article stating that "Vänsterpartiet (the leftmost party in Sweden) is going to crash the economy through their reckless spending on public projects" that would be biased since the words used have negative connotations. Although this example is perhaps a bit tame and there are several more direct ways to write damning statements all the more it is important to recognize that bias comes in many forms, be they subtle or not. Bias often becomes more visible when two separate subjects are put in juxtaposition to each other. A further example would be to follow up the aforementioned article damning Vänsterpartiet with something praising Moderaterna or Liberalerna for their responsible cut backs to public spending. A clear preference is established between the alternatives and therefore a bias is present.

In this project two types of bias are investigated: visibility bias which is the amount of coverage given to any given party and sentiment or tonality bias which is the tone or sentiment which the newspaper hold towards a given party.

1.1 Purpose and Goals

This project attempted to categorize bias in online newspapers as well as evaluate the methods used in doing so. Two types of biases were investigated: visibility bias and sentiment bias or tonality bias.
1.2 Bias

Visibility Bias: Visibility bias is quite self-explanatory, the amount of exposure given to a party or party representative. There is a popular saying that “All publicity is good publicity” and if there ever were a way of testing that statement visibility bias is that way.

Tonality Bias or sentiment bias: Whilst sentiment bias is quite easy to explain it is significantly harder to measure. In essence sentiment bias or tonality bias is all about the tone or meaning behind a set of words. In this paper a ternary definition is used where a statement can be either positive, neutral or negative. Another tricky part of sentiment analysis is how to determine what a statement is biased towards [2]. That is further discussed in the model part.

Agenda Bias: Agenda bias is perhaps the hardest to grasp, it is similar to visibility bias since in essence is about what gets shown or hidden. Imagine a newspaper each day get hundreds of stories. Among them only a subset gets published, agenda bias is the bias the editors exhibit when they choose which stories are to be published. However measuring this would require access to all discarded stories which is quite difficult [3, 4].

These sub-biases were defined by Markus Wagner, Hajo G. Boomgaarden and Jakob-Moritz Eberl in their paper One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences. However, these definitions are not entirely unique as similar systems have been used before [4].

This paper only contains two biases, visibility and tonality. This is due to the impracticality to implement agenda bias in any meaningful way without access to each newspapers discarded stories. Since there is a form of overlap between visibility bias and agenda bias and the relative difficulty of data collection for agenda bias it was decision made to reduce the scope of the project.

1.3 Visibility Bias

In the case of visibility bias the approach is much the same as in the aforementioned paper “One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences”. A party is considered visible in the article in question, if there is any mention of it or the representative for that party. However, as Hopman argues visibility bias should be compared to benchmarks of parties at a given time. For example their relative amount of seats in the parliament. Furthermore, Wagner argues that since news outlets could simply be less politically inclined and in general leave less space for politics in their issues. It is also important to include the average visibility of each party for each
newspaper. To make an example: If any given party A is visible in 60% of all politically inclined articles from any given source B and for that same source the average visibility is 40%. Meaning that on average each political party is mentioned in 40% of the politically inclined articles. That means that party A in newspaper B is given a visibility bias of 20 percentage points. This insures that different newspapers are comparable despite differing ratios of political news.

1.4 Tonality Bias

Tonality bias or sentiment analysis have long been used by companies to get an overview of how the public and or the customers view their product. For example business analytics have been data mined to extract patterns to better improve marketing strategies [2].

Once again similarly to the paper “One Bias Fits All? Three Types of Media Bias and Their Effects on Party Preferences” tonality bias is measured by aggregating statements about parties and their representatives. These statements are graded on a scale from -1 to 1 with -1 being extremely negative and 1 being extremely positive. However since this is achieved using sentiment analysis and not manual content analyses the similarities end there. Furthermore, since Vader was the tool employed (Vader is further discussed in section 2 and alternative approaches in section 3) which also uses a tertiary system with a neutral score of 0. It is technically possible for a newspaper to be completely unbiased considering the vast amount of articles analyzed it is exceedingly unlikely, it is also worth mentioning that even if individually articles might be biased the average might still become unbiased. If this is the case it is unfortunately not covered in this project since only the averages are measured.

In essence sentiment analysis is in general applied on three different levels; document, sentence and aspect. Farhadloo and Rolland argue that a document or article in this case can contain an overall negative or positive sentiment but since a document or article can contain points and/or counterpoints it is important to extract the relevant data [2]. Applying this in the context of this project if an article mentions a political party or representative it is safe to assume that the article is in some way political. However, if one were to simply extract the sentiment on a document level and apply to all mentioned parties the complexity of the article would most likely be lost. Therefore, a sentence level approach is necessary because an article might be negative about party A’s approach to a certain issue it could also praise party B for their approach to the same issue.

Moreover, there are three types of sentiment analysis; unsupervised models, supervised models and lexical models. This project uses a lexical based approach but there is merit to all of them.
Unsupervised models use two seed terms which for the English language would be “Excellent” and “Poor”, for each unknown word the similarity between the word and the two seed terms are then computed using a search engine to count the occurrences of the two seed terms respectively. From this the algorithm then determines the leaning of the word. However, for this to work properly one must first extract the opinionated fragments and the overall topic or aspect of a given text. According to Farhadloo and Rolland it is sensible conjecture that adjective, adverbs and also some nouns and verbs are strong sentiment indicators. These words can then be marked for the unsupervised algorithm.

Supervised models use the fact that it is relatively easy to make any given sentiment identification problem into a classification problem. This is done using machine learning to identify the sentiment of a large unidentified set of data from a small identified set of data. Naive Bayes classifiers are typically used to train the classifier on the small set to then identify the larger set. Another approach is to use heuristic deep learning to both learn the sentiment and aspect of a given document at the same time. This can be done by creating a parse tree for any given phrase in which the nodes are represented by a vector and the model takes advantage of the syntactic structure of the phrase.

Lexical models which is what is being used in this project, determine the sentiment of any given statement by looking up the phrase in the given lexicon. However, additional rules are applied for negations such as “not”, ”but”, “neither” etc. This is also true for connective words such as “and”, “or”, etc [2]. More on this in section 4.4

Vader as a tool was chosen due its accessibility and due to the fact that Vader is a lexical system a training set was not needed. Since this project did not have access to a group of experts which could categorize individual articles bias and thus create a suitable training set for a supervised model. Such an undertaking would also fall outside both the scope of the project and allotted time however an interdisciplinary project to create a training set and subsequent classifier is an interesting avenue for future research.

2 Related Work

Mapping political bias of newspapers is not a new concept and it has been done several different ways before. For example as previously mentioned Markus Wagner, Hajo G. Boomgaarden and Jakob-Moritz Eberl who evaluated the data from the Austrian parliamentary election campaign of 2013, in the study they map press-releases by political
parties to the media coverage of those press-releases. More specifically they used the data from the Austrian National election study (AUSNET) which they then compared an online survey that measured in four waves how the Austrian populations party preference shifted [3]. What makes this interesting in relation to this project is the similarity between the Austrian parliament and the Swedish parliament. Both use a multi-party system which according to the aforementioned paper is rarely analyzed, a notion that I personally agree with after looking for related works. Moreover, similar to how this project was conducted they also used analysis on a sentence level. However, one key difference is the time span and amount of sources investigated. They only looked at a time span of six weeks between August 19 and September 29 (2013) for 8 newspapers whereas this project analyses 8 months of data from January 1 to September 10 (2018) for 4 newspapers. Curiously enough the subset of political articles extracted is similar in size suggesting the increased intensity coverage closer to election.

This project uses sentiment analysis to extract the tonality bias of the different parties which in of itself is a relatively new field with many applications the chapter Fundamentals of Sentiment Analysis and Its Applications by Mohsen Farhadloo and Erik Rolland from the book Sentiment Analysis and Ontology Engineering was of great help in both understanding and applying the methods used in this project. According to Farhadloo and Rolland there are two ways to quantify bias, a binary representation of positive or negative is most common. However, some add a third neutral sentiment making it tertiary system. However a sentence might be neutral but a whole article or in the case of the chapter document is highly unlikely to actually be completely neutral.

Furthermore, the implementation of the sentiment analysis in this project uses the Swedish version of the tool Vader (Valence Aware Dictionary for sEntiment Reasoning) sentiment analysis for python which uses a “generalizable, valence-based, human-curated gold standard sentiment lexicon” [5]. This in turn means that Vader does not require any training data since it pulls the valence values from the lexicon. Despite being a lexical tool it still outperformed human testers when testing tweets in F1 score which is a measure of a test’s accuracy. Vader considers both the precision and the recall of the test to compute the score. However when editorial pieces were considered Vader did not outperform human testers [5]. While Vader itself was made by C.J. Hutto and Eric Gilbert and is discussed in high detail in their paper VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text the Swedish version is maintained by Alex Gustafsson. Whether it is just the lexical part that he maintains or the code is hard to gauge from the pypi.org and github.com. In a way this project will serve as a test run for the Swedish port of Vader for sentiments on articles from newspapers.

Speaking of the Swedish Vader port it was made by Marcus Gustafsson [6] and explored in further detail in his thesis Sentiment Analysis for Tweets in Swedish. Although the name Alex Gustafsson is never mentioned there is a possibility that it is a pseudonym under which he uploaded the code. For this project the relevant part of the thesis is the
fact that the aforementioned "gold-standard" sentiment lexicon was machine-translated to Swedish which inevitably lead to a reduced F1 score [6].

3 Background

3.1 Previous efforts

The original plan of the project was to do data gathering during the duration of the project by gathering data from a predetermined set of online newspapers. This was to be done by connecting to the respective rss-feed of each newspaper. However since the primary focus of this project is to evaluate the models and methods used. To achieve this it was deemed more important that the data used should be as clear as possible. Consequently, it quickly was replaced by suggestion of the project supervisor. Gdelt was the second chosen approach which contains machine-translated data from a plethora of Swedish sources [7]. Gdelt proved to be quite a challenge for me personally due to my inexperience with google big Queries which seemed like the easiest way to narrow the scope of the staggering amount of information available at Gdelt. Alternatively one could download the raw data from Gdelt and then sift through that, however both the hard drives and internet capacity available to me made that a non option. Lastly once again by recommendation of the project supervisor Mediacloud was employed to gather the necessary data. Mediacloud has both a web-based service as well as an api. [8]Firstly, the api was tried to directly access the data but much like inexperience was my downfall with gdelt the opposite was true here. Feeling confident in my ability to handle the api I barely spared the website a glance before diving in to the documentation. Thankfully I discovered relatively quickly that extraction of source links was trivially easy through the web portal.

Backtracking to the original plan once the data was collected the plan was to use Rapidminer to extract the bias from the text. However, since most of the refining of the text as well as the scraping for the raw text already took place in python it felt more natural for me to continue working exclusively in python.

So in the end every facet of the original plan except the overall purpose got changed.
4 Methodology

4.1 Source selection

The sources were selected to have a wide range of purported leanings, the leanings were gathered from Alexandra Segerberg who is senior lecturer at the Department of government in Uppsala. These leanings were also used to serve as a control for the results.

Expressen: Tabloid with liberal leaning

Aftonbladet: Tabloid with self described leaning of “unbound socialdemocratic”

Svd: Edited on a value basis of united liberalism and conservatism

Svt: Public service broadcasting

4.2 Data Gathering

The data collection was done using mediacloud which offers both an api-service and a web portal. The web portal allows the user to access links to date-marked articles from as far back as 2011. Moreover, different sources have different starting dates. For this project links from the first of January to the tenth of September 2018 were used from the sources: Expressen, Aftonbladet, Svd and Svt.

The dates were chosen since 2018 was the most recent election year and should therefore have a high saturation of political articles. The dates were also chosen with the same objective since there is no use comparing the bias of the newspapers after the election results. Therefore the dates are from the beginning of the year to the day after the election.

Once the links were downloaded they were exported to a txt file where python script used them as the basis for a web scraping script which made the raw text available for sentiment analysis as well as visibility analysis. From all the sources handled, an average of 0.7% were discovered to be broken links no longer available.
4.3 Tonality/Sentiment bias

One of the harder parts about sentiment analysis is to determine the relevant context or aspect of any given article [9]. Due to this in the model used for this project it was deemed more appropriate to use a catch all method where each sentence was checked for keywords which corresponded to any given Swedish party. For each party mentioned the sentence was added into evaluation for that party, for example if a sentence mentions both Party A and B then that sentence would be evaluated for sentiment bias for both parties. Furthermore, all subsequent sentences would be added as “Context sentences” for all aforementioned parties until another relevant keyword was detected at which point the sequence starts over.

As illustrated in figure 1 above all articles are broken down into political articles by isolating all articles that mentions any political party or their representative. These political articles are then divided into political sentences and context sentences as in the example above. The political sentences and context sentences are then grouped together into a sentence group which is then evaluated using the lexical Vader system. Lastly the sentiments for each party are then divided by the number of groups for each party which gives the sentiment or tonality bias for that party. However, it is important to note that figure 1 is just an example of how one sentence group is evaluated, articles typically have several groups.

As previously mentioned the method used for the sentiment analyses is based on the

<table>
<thead>
<tr>
<th>Newspaper</th>
<th>Nr articles</th>
<th>Nr political articles</th>
<th>% of political articles</th>
<th>Broken links</th>
<th>% Broken links</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD</td>
<td>28325</td>
<td>2826</td>
<td>10%</td>
<td>81</td>
<td>0.29%</td>
</tr>
<tr>
<td>SVT</td>
<td>39761</td>
<td>2668</td>
<td>6.7%</td>
<td>553</td>
<td>1.39%</td>
</tr>
<tr>
<td>Aftonbladet</td>
<td>22735</td>
<td>1534</td>
<td>6.7%</td>
<td>204</td>
<td>0.9%</td>
</tr>
<tr>
<td>Expressen</td>
<td>30420</td>
<td>2404</td>
<td>7.9%</td>
<td>64</td>
<td>0.21%</td>
</tr>
</tbody>
</table>

Table 1: Spread of articles for each newspaper
Vader system for python with the Swedish translation maintained by Alex Gustafsson [10]. Vader is a lexical based sentiment analysis tool. There exists other alternatives such as the Linguistic Inquiry and Word Count (LIWC, pronounced “Luke”). It was originally intended for social media use but has seen use beyond that scope. One very tangible advantage with Vader over LIWC is the ability to discern the intensity in words for example in LIWC 'okay' and 'exceptional' would carry the same weight in a sentence (one positive word) whereas Vader would rank 'exceptional' higher since it is a more intensely positive word [5].

Vader uses a tertiary system which means that it rates sentences from -1 to 1 where -1 is extremely negative 0 is neutral and 1 extremely positive. Three aspects are measured: neg, neu and pos which simply measures how neutral, negative and positive a statement is[10] These three aspects together form the compound score which is what was used in this project.

For a compound score > 0.05 the statement is considered positive.

For a compound score < -0.05 the statement is considered negative.

Lastly for a compound score in the interval 0.05 < x > -0.05 is considered neutral [10].

As previously stated the Swedish port of Vader was machine-translated which means that some underlying complexity is lost [6]. Furthermore, if one considers the relative complexities between the languages which in of it self is a hard question to answer then it becomes natural that some underlying complexity is lost in translation. According to Oxford dictionary the English language may be the language in the word with the most words [11]. This assertion does however presuppose that so called connective or conjunction words are ignored. For example Nordvästersjökustartillerifygspaningssimulatoranläggningsmaterielunderhållsuppföljningssystemdiskussionsinläggsförberedelsearbeten is technically a word in Swedish although it is not accounted for in this comparison.

This becomes the most apparent when looking at inflexion words, which in turn means that words like 'fine', 'good' and 'great' all get the same translation "bra" although the all have different valence scores in the original lexicon. The consequence of this is that three separate valences values exists for the word bra [6]. The solution to this was to use the average but of course some complexity was lost by this imperfect solution.

4.4 Visibility Bias

The visibility bias is computed in a much more straightforward manner. Once again first the political articles are extracted from all the articles. (This does in fact not happen twice in the code but it helps to visualize.) The political articles are then traversed and
each time a party or its representative is mentioned, that article is counted towards the visibility of the party in question. All the visibility biases are then accumulated into the average visibility for that newspaper. See figure 2 below for an overview of how the process works.

Figure 2: How visibility bias gets computed

It is important to note that whilst the sentiment analysis was conducted on a sentence level the visibility bias is conducted on an article level. Since amount of mentions per sentence is not really a statistic that is significant from a bias standpoint. Although it is important to acknowledge if a party is continuously mentioned several times throughout an article that parties actual visibility bias would be higher than the actual bias displayed.
5 Result

5.1 Visibility Bias

Although visibility bias is not necessarily a measure of whether or not a given newspaper is inclined towards one party or another, since a very politically opinionated newspaper could simply only report negative news of any given party. However, it is still a valuable metric to get an overview of which parties has the most interest, positive or negative.

![Figure 3: Percent of votes from the election before 2018](image)

For reference figure 3 contains the percent of votes each major party had in the 2014 election.

5.2 Visibility Bias Svd

As seen below in figure 4 Svd reports quite evenly across the smaller parties while the two biggest parties take the forefront. Interestingly enough Svd is the only newspaper in this study that reported more on the second largest party Moderaterna. This could however be due to the fact that according Alexandra Segerberg Svd as a newspaper is...
Figure 4: Visibility bias Svd

described as "unbound Moderat, edited with liberal and conservative values" [13].
5.3 Visibility Bias Aftonbladet

Aftonbladet showed some small fluctuations across the smaller parties with the notable example of Sverigedemokraterna. Although, since Aftonbladet is a tabloid which are known to report on more sensationalist topics [13]. Tabloid sensationalism could be a factor as to why Aftonbladet gave more coverage to Sverigedemokraterna since a lot of their policies generate a lot of controversy [14].

A comparable disparity can be found in the amount of coverage given for Kristdemokraterna as well as Vänsterpartiet despite both of these being at opposite sides of the political spectrum. Both parties were quite small after the 2014 election as seen in figure 3. According to the control provided by Alexandra Segerberg Aftonbladet is described as "unbound social-democratic"[13] which is reflected in the coverage.
5.4 Visibility Bias Expressen

Expressen also showed some fluctuations in the amount of coverage given to the smaller parties. Moreover much like Aftonbladet, Expressen is also classified by Alexandra Segerberg as a "liberal" tabloid. Interestingly enough, despite the fact that Expressen is categorized as "liberal" there was no significant difference between Expressen and the other sources on the amount of coverage given to the liberal party (Liberalerna). Lastly Expressen shows some bias towards Sverigedemokraterna but not as much as the previously discussed tabloid Aftonbladet. Which contradicts the aforementioned theory of why Aftonbladet would report more on Sverigedemokraterna due to the controversial nature of the party.

![Visibility bias Expressen](image)

**Figure 6:** Visibility bias Expressen
5.5 Visibility Bias Svt

Svt is an interesting case, as mentioned in section 4.1 Svt is a public service news site as such it should be as objective as possible. Interestingly enough when you consider the relative sizes of the parties from the 2014 election nothing really stands out except for perhaps a bit of overexposure for Sverigedemokraterna. Although since there only is one newspaper that haven’t covered Sverigedemokraterna thoroughly(Svd) a case could be made that Svd is in fact biased by under-representing the party.

![Figure 7: Visibility bias Svt](image_url)
5.6 Sentiment/Tonality bias

As previously mentioned in section 4.3: This is how the Vader system compound score works.

For a compound score > 0.05 the statement is considered positive. For a compound score < -0.05 the statement is considered negative. Lastly a compound score in the interval 0.05 < x > -0.05 is considered neutral [10].

The most interesting results is that not a single newspaper is on average negatively inclined towards any party. Expressen and Svd are almost completely neutral. Whilst Aftonbladet and Svt are either neutral or positive. Aftonbladet is a newspaper with a "Social-democratic" leaning [13]. However the party that Aftonbladet is the most positively biased towards is Vänsterpartiet. Both Vänsterpartiet and Socialdemokraterna is part of the same political block so the shift in from the expected leaning is not big but still relevant.

Furthermore it must be stated that these results are quite inconsistent with what was expected. Considering that both Expressen and Aftonbladet are both tabloids[13] they were expected to capitalize on more sensationalistic type of reporting and therefore exhibit more of a bias whereas Svt which is a public service news page was expected be the least biased in this group. However this does not correspond with the results shown above. It should also once again be noted that none of the newspapers appeared in the negative range. This could be an issue with the translation mentioned in section 2 or perhaps it is simply how large Swedish newspapers operate. It could be interesting to evaluate the sentiment of entire articles for a larger sample size of newspapers to test whether most Swedish newspapers are in fact positive. It could also be due to the large sample size: when the algorithm was tested on individual articles the bias values returned were significantly higher. It could also be due to the fact that the model can
evaluate several parties at the same time. As explained in section 4.3 if a sentence contains references to multiple parties subsequent sentences will be added as context to all referenced parties. Therefore if that sort of sentence is common the results will become more uniform. However whilst the total range is quite small, variation between the newspapers still exist within that range consequently it seems unlikely that this is the case.

6 Conclusion

During this project an attempt was made to categorize bias in online newspapers, using visibility bias and tonality or sentiment bias to do so. Both approaches were naive in their implementation and served as a test for the models described in section 4.

The visibility bias yielded interesting results with only three parties (Moderaterna, Socialdemokraterna and Sverigedemokraterna) consistently being reported on above the average visibility for each site. Which also are the three biggest parties within the Swedish parliament. However only one of these (Sverigedemokraterna) actually gained in mandates after the election[12]. Consequently drawing any conclusion on whether or not the increased exposure increased voter support is tenuous at best. Although Sverigedemokraterna did gain a sizeable amount of support it could also be attributed to the general increase in support for right-wing populist parties during the same period in Europe[15]. Furthermore, all small parties except Miljöpartiet actually increased their mandates during the 2018 election [12]. This despite Miljöpartiet being one of the most visible of the smaller parties except Sverigedemokraterna, although with small margins. In the end it is difficult to say whether or not all pr is in fact good pr. Historically speaking newspapers have been hard to quantify bias-wise[16]

Sentiment analysis have historically struggled when applied to editorial pieces [5][2] and that is seemingly true for this project as well. The project was designed to make use of a scale from -1 to 1 but only the interval 0 to 0.16 was necessary, there could be numerous reasons for this which are explored in section 4.6. However in essence it is once again difficult to draw any concrete conclusion since the data was inconclusive.

6.1 Future Work

I personally think that the future for projects similar to this one in Swedish are huge. There already exists a staggering amount of research for the same concepts in English. Simply transferring that knowledge into Swedish application could yield fascinating results. I think this type of work is best suited for an interdisciplinary group with both knowledge of the subject in question (in this case politics) and the underlying methods.
Most likely a linguist could also be of enormous use, at least if a lexical approach were to be attempted. Personally I struggled a lot with coming up with suitable models on how to approach the problem in the first place.

To whomever might want to do a similar project I can wholeheartedly recommend the visibility bias approach used in this project. I think a naive approach is best suited for such bias due to the simplicity in categorizing text by content simply by using keywords or names. It might also be prudent to include some consideration for pictures as they are completely ignored in this project. A possible improvement could also be to rank the articles by number of mentions within the article itself perhaps visibility bias should be conducted at a sentence level similar to sentiment bias.

When it comes to sentiment analysis I believe that a better approach would have been to go with machine learning. There are several models that can be applied to the Swedish language such as Naive-Bayes and support vector machines (SVM) both of which have been used successfully before in regards to sentiment analysis [2]. Furthermore, this would probably become an interdisciplinary issue since a training set would have to be formed with preset biases which most likely would require someone well versed in political science or journalism. Moreover, research shows that regardless of training model the training-sets are actually more important for machine learning [17]. It also bears to mention that Vader have proven at least in the English language that lexical rule-based systems can still compete with machine learning and humans. However, to adapt this to Swedish use I believe that it once again becomes an interdisciplinary issue to properly translate the lexicon as well as rewriting the syntactic rules.
7 Figures

Figure 8: Percent of votes from the election before 2018 [12]
Figure 9: Visibility bias Svd

Figure 10: Visibility bias Aftonbladet
Figure 11: Visibility bias Expressen

Figure 12: Visibility bias Svt
References


