Navigating Politics: A NLP Exploration of Danish Parliamentary Debates

Freya Rebekka Foldager Lauritsen (FRFL) (202207040@post.au.dk)

Magnus Severin Ringgaard Poulsen (MSRP) (202208478@post.au.dk)

School of Communication and Cognition, University of Aarhus,

Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Social and Cultural Dynamics

Supervisor: Arnault-Quentin Vermillet

03.06.24

Abstract {FRFL & MSRP}

This paper aims to investigate how topic modeling and sentiment analysis can be utilized in exploring tendencies in the Danish parliament. In doing so, exploring how topics have evolved, how they are distributed between blocs, and how single events, such as the climate election, can affect these trends. Furthermore, how being in government or opposition affects the topics and tone, and lastly the influence of gender and age on the topics discussed. This is done by first outlining the Danish political landscape and how demographics influence political positions, and introducing topic modeling and sentiment analysis in the light of parliamentary debates. Secondly, topic modeling and sentiment analysis applicability on Danish parliamentary debates is investigated by analyzing a corpus of 1704 transcripts. The findings suggest that topic modeling and sentiment analysis are useful tools for detecting tendencies in parliamentary debates. Topic modeling is found to be able to reflect several events in society. In addition, it can be used to reveal tendencies at various levels; time, gender, age, topic, bloc and party - by combining these it is possible to explore tendencies that interact, thus yielding more comprehensive findings. Sentiment analysis was likewise found to be used for detecting tendencies, as well as interplay with other findings using topic modeling, creating more comprehensive results. Lastly, limitations concerning classification, party change, and sentiment analysis are discussed.

Keywords: topic modeling, sentiment analysis, BERTopic, Danish parliamentary debates

Introduction {MSRP}

Politicians communicate to voters through many channels; public speeches, social media and legacy news. Through these, politicians can shape their image towards the public, and this will undoubtedly influence communication. However, this is not where their actual work happens. It happens within the Danish parliament, and the meetings held there give the clearest insight available to the public into the workings of Danish politics. Transcripts from the parliamentary chambers (Folketingssalen) are accessible from the Danish parliament's website ranging back to 2005, but sifting through reading all these documents is a laborious task. However, with the tools provided by natural language processing (NLP), this becomes less daunting. This paper will serve as an exploratory investigation aiming to elicit how NLP tools can be used to investigate underlying trends and biases in Danish politics by analyzing 1704 transcripts of parliamentary debates from 2008-2023.

The Danish Political Landscape {FRFL}

To understand the contents of the debates in the Danish parliament it is crucial to understand the current Danish political landscape. The following section will outline current political trends and the foundations of Danish politics.

Denmark's political system is a representative democracy consisting of 179 parliamentary seats distributed among 11 political parties and two mandates each for Greenland and the Faroe Islands. Parties are organized in a bloc structure, where left-leaning parties are grouped in the red bloc with Socialdemokratet (*The Social Democratic Party*) being the largest and leading party, and right-leaning parties organized in blue bloc historically led by Venstre (*The Liberal Party*) (Skidmore-Hess, 2003).

Over time the Danish political landscape has shifted, and so have politicians. A study investigating party changing from 1953 to 2015 found a significant increase, and since this trend has become even clearer, with 18% of politicians changing party between the election in 2019 and 2022 (Toft, 2022). This is partly an effect of the number of parties in the Danish Parliament increasing since 2000, and following this parties' mandate counts converging, becoming more evenly distributed.

Research by Green-Pedersen & Kosiara-Pedersen (2020) suggests the increasing number of parties being linked with a development in the topics discussed in the parliament in recent years. Research by Green-Pedersen (2013) and Skjæveland (2005) finds this development in topics shifting from centering around economy and redistribution policy to topics regarding the environment and immigration. On one side due to the increasing importance of the refugee and immigration issues, this is suggested to have led to the establishment of multiple parties profiling themself on anti-immigration; Dansk Folkeparti (*Danish People's Party*), Nye Borgerlige (*New Right*) and Danmarksdemokraterne (*Denmark Democrats*). On the other hand, the increasing importance of the environmental topic has led to the establishment of Alternativet (*The Alternative*) and Frie Grønne (*Free Green*, now disbanded).

With the Danish political scene developing both in terms of parties and topics, so does voter behaviour. Hansen & Stubager (2024) found that in the latest Danish general election in 2022, voters displayed a general inconsistency with their previous voting pattern, seen in 53% of voters changing party - a record for voter migration in Denmark. It is suggested this could be due to politicians' personal image becoming increasingly important for voters when deciding which candidate they prefer. Numerous studies argued that social media strongly reinforces this personalization trend (G.

Enli, 2015; G. Enli & Moe, 2013; G. S. Enli & Skogerbø, 2013; Severin-Nielsen, 2024; Skogerbø & Larsson, 2021). Social media platforms such as Facebook, Instagram, X, and TikTok make it possible for politicians to define their image even more compared to legacy news media (G. Enli & Moe, 2013; G. S. Enli & Skogerbø, 2013; Severin-Nielsen, 2024). In 2020 only three out of 179 Danish parliament members did not have a public page on either Facebook, Twitter, or Instagram (Severin-Nielsen, 2024). This strong social media presence coupled with voters' focus on politicians' image, whilst knowing social media appearances can be misleading (Tucker et al., 2018), makes it crucial to ascertain the actual work being done in the parliament.

Demographics {MSRP}

When investigating parliamentary debates party and bloc affiliations should not be the only points of interest, especially taking the tendency of party change into account. Exploring demographic differences such as gender and age is also valuable, especially as these play an increasingly significant role in Danish voters' political stance. This is outlined in the work by K. M. Hansen (2019), in a review of the demography of Danish voters in the election of 2019. It is found that the discrepancy between genders voting for the red bloc has grown since the 1980's, where there was only a 3% difference in votes coming from women versus men, as opposed to having increased to 10-11% at the 2015 election (K. M. Hansen & Stubager, 2017). Furthermore, men are increasingly voting to the right and women to the left. A prognosis for the 2019 election found that the majority of voters for Socialistisk Folkeparti (*Socialist People's party*) were women, constituting 60% of votes, whereas far-right parties such as Stram Kurs (*Hard Line*, now disbanded) and Nye Borgerlige mostly attract men, with them making up 75% of votes. K. M. Hansen (2019) suggests that this discrepancy can be attributed to women valuing welfare higher than men over the past 20 years, with a growing divergence in views on economic inequality and redistribution. Returning to politicians, female politicians hold 42% of the parliamentary seats in 2023 with women holding 52% of the seats in red bloc, and 40% in blue bloc (*Mandatfordelingen*, 2024). In addition, studies of transcripts from the British Parliament suggest that gender also influences which topics are addressed. Blaxill & Beelen (2016) found that female members of parliament speak more frequently about topics regarding children, healthcare and education, whilst male members speak more of business, foreign affairs and defence.

Another important factor suggested by K. M. Hansen (2019) is age influencing political orientation. It was found that together, Socialdemokratiet and Venstre had 54,2% of the votes from the elderly above 60 years old in 2019. In the category 19-29 years old, their share of votes is halved. In the category of 30-59 years old, the votes are spread more evenly across all parties. Age can strongly influence which topics matter for voters, for example, 40% of young voters in 2019 cast their vote in favour of the four parties with the highest ambitions regarding climate and environment, though only receiving 26% of the total votes. Research has also shown general age differences such as young voters favoring investments in the environment and older voters supporting higher pensions (Metz, 2002; Stockemer & Sundström, 2018). Thus age seems influential in political stance, and as such investigating this along with gender seems important for understanding politicians' political stance, and whether this is influenced by demographic traits the same as voters seem to be. K. M. Hansens' research also points to education as an important factor for voters' decision making, but will not be explored in the current analysis, due to politicians' relatively uniform educational background, with 73% having a degree or higher education (Nørtoft, 2019).

Natural Language Processing and Parliamentary Texts {FRFL}

To explore the above-mentioned areas of interest in the large corpus of parliamentary transcripts, tools from NLP will be employed. NLP is a research area, where machines are utilized to process and generate language. The goal is to mimic human-like processing for a variety of tasks such as topic modeling and sentiment analysis (Louis, 2020, para. 1). In the context of analyzing parliamentary speeches and debates, NLP as a method has widely been applied (Abrami et al., 2022; Hyvönen, 2022; Navarretta & Haltrup Hansen, 2020; Perak & Rodik, 2018).

Topic modeling is the process of categorizing texts, referred to as documents in NLP terminology, that hold similar meaning, to uncover shared semantic structures, thereby giving an insight into the themes or general topics in a corpus of documents (Dahal et al., 2019; Umamaheswaran et al., 2023). The generated topics are not given based on any prior categorization from a human. Instead, topics are classified by advanced machine learning algorithms, which will be expanded upon in the following section. This process is known as unsupervised learning and is useful when analyzing large amounts of raw data for topics and patterns. As topic modeling only acts as a surrogate for human labeling of large corpora, it is common practice for topics to be reviewed and often further manually clustered (D. H. Hansen et al., 2019).

A study investigated the use of NLP to explore parliamentary debates in the French Third Republic from 1881-1899 (Bourgeois et al., 2022). The study found that topic modeling is a useful tool in analyzing parliamentary debates. In addition, it found that parliaments follow their own rhythm imposed by legislative process, which requires long debates before becoming newsworthy. Opposed to this, issues in the news are rarely discussed during parliamentary meetings. Usually, they are dealt with when they are out of the news cycle. Therefore, topic modeling seems to enable the identification of underlying political trends, aligning with the scope of this paper.

Besides topic modeling, sentiment analysis will be used for a more in-depth investigation in this paper. Sentiment analysis is used for identifying and extracting information about the emotional valence of a document (Wankhade et al., 2022). This is often extracted to a polarity measure, classifying whether documents are positive, negative or neutral (Taboada, 2016). The possibilities this measure offers are great in the scope of analyzing parliamentary debates and have been used to infer; agreement and alignment, ideology and party affiliation, and scaling of political position (Abercrombie & Batista-Navarro, 2020).

BERTopic and Parliamentary Texts {MSRP}

Previously topic modeling has been done mostly using Latent Dirichlet Allocation (LDA) models, which, in crude terms, estimate the probability that a word occurs in a topic given its frequency and the probability of a topic occurring in a given document. However, with the emergence of neural network-based transformer models, such as Google's BERT, pre-trained on large corpora - topic modeling can now utilize information not only from single words but also their context, allowing a deeper semantic embedding (Kozbagarov et al., 2021). The Python package, BERTopic, leverages the infrastructure of BERT to elicit coherent topics from documents (Grootendorst, 2022). The algorithm works using a pipeline consisting of; document embedding, dimensionality reduction, and creation of topic clusters and representation. The process of document embedding is done by a pre-trained sentence-transformer model, which converts documents into numerical representations. These

embeddings' dimensionality are then reduced using UMAP to be able to cluster them. Next, documents are clustered from their embeddings, and finally, topic representations are created, using the c-TF-IDF algorithm, which compares word importance by their frequency between clusters. Topic representations are displayed as a set of words best describing a topic cluster and a list of representative documents. A feature of BERTopic is that it is modular, so that the steps of this pipeline can be customized, so for example, document embedding can be done using different pre-trained transformer models.

A study by Egger & Yu (2022) compared the topic modeling methods LDA, NMF, Top2Vec and BERTopic, finding that Top2Vec and BERTopic outperformed the others due to their use of embeddings, resulting in more meaningful and coherent topics. Further studies have shown BERTopic outperforming other models in classifying political text including parliamentary speeches (Contreras et al., 2022; Ivanusch et al., 2022; Lazreg, 2023; LI et al., 2024).

In the context of Danish Politics, Navarretta & Hansen (2023) examined the policy areas "environment" and "energy" by comparing seven red and blue bloc parties in their electoral manifestos and parliamentary debates between 2009-2020, finding BERTopic useful for extracting political issues in parliamentary speeches about specific policy areas. Additionally, they compared two BERTopic models; Danish and multilingual, finding multilingual BERTopic generated more relevant clusters. Furthermore, the paper found that the parties' frequency at which they addressed the topics "environment" and "energy" differed in their voter-directed manifestos compared to their parliamentary debates. This again supports the importance of investigating the underlying political trends in parliamentary debates, as they can differ from politicians' or parties' explicit voter-directed communication.

Classification of Political Topics

As mentioned after topic modeling, the resulting topics are often further manually clustered. In the realm of political topics different systems have been developed with the purpose of categorizing political topics. The Comparative Manifesto Project, CMA, and Comparative Agendas Projects, CAP, have developed two classification systems for comparative studies (Budge et al., 2023; Loftis & Mortensen, 2020). In the context of Danish politics a study by D. H. Hansen et al. (2019) exploring how to cluster data from Danish parliamentary debates, found that these systems are too complex and broad to classify Danish parliamentary debates. Based on that, they created a "gold standard corpus for automatic classification''. The study used 22 topics proposed by Danish scholars in political sciences for their annotation work. The relevant areas of the speeches were manually annotated, assigning the subject areas to the corresponding speeches. They found that the gold-standard corpus can be utilized to automatically classify Danish parliamentary transcripts with a high accuracy. The following analysis will use the topics from this work.

Research Question {FRFL}

The large dataset of transcripts from Danish parliamentary debates and the emergence of NLP with its possibilities of topic modeling and sentiment analysis, makes it possible to investigate underlying trends within Danish politics exploring what is actually happening inside Christiansborg, rather than exploring what the politicians choose to present directly to the public. With the current shifting political landscape it seems relevant to investigate how parties react. Further demographic information

gender and age make for interesting points of analysis, as they seem to impact political stance. On this basis, this paper serves as an exploratory paper aiming to investigate how topic modeling can be used as a tool for exploring political tendencies in parliamentary debates. This is posed as the following research question;

RQ: How can topic modeling and sentiment analysis be used as tools in exploring Danish parliamentary debates?

Aiming to demonstrate the diverse applications of natural language processing, this study addresses the following subquestions;

1. What are the main trends of Danish parliamentary debates and how have these evolved? In doing so, investigating how topics have developed over time, and how they are distributed between blocs. Following this, how can single events, like the Climate Election of 2019, explicate these trends? Furthermore, how has the tone of parliamentary debates evolved?

2. How does being in government or opposition affect topics and tone?

3. What are the influences of gender and age on the topics discussed in parliamentary debates?

Methodology and Results

Data Acquisition {MSRP}

Parliament transcripts are freely accessible from the Danish Parliament's website (*Referater*, 2016), and were downloaded along with information about members of parliament, also accessible through the website (*Oda.ft.dk*, 2024). Transcripts contain meeting agendas, voting outcomes and importantly for the current purposes; text records of everything uttered in Folketingssalen (*parliament chambers)*. Records range back to the 19th of April 2005 and are continuously being updated. A change in document formatting after the 3rd of October 2007 led to data before this time being left out for analysis. Furthermore, for the simplicity of the analysis data from incomplete years (2007 and 2024) were also omitted, resulting in the current analysis concerning data from January 2008 to December 2023. In this time frame 1704 transcripts were recorded.

Concerning information about members of parliament, the collected metrics were; full name, age, gender[1], and latest party. A total of 462 current and former members of parliament spoke in the collected transcripts.

Data Preprocessing {FRFL}

The transcripts were split into segments by speaker, such that every time a new speaker spoke this would be saved as a document. This document extraction was done in Python (Van Rossum & Drake, 2009) using BeautifulSoup (Richardson, 2007). Further preprocessing consisted of the following steps: (1) extracting the speakers name for each document, (2) removing documents from the Speaker of the Parlimentary Chambers, as their job consists of mediation of the debates, and as such almost

[1] The Danish Parliament's records report politicians sex and this will be interpreted as gender.

exclusively consisted of short parliamentary procedure formalities, (3) removing one-word parliament procedure formalities[2], greatly decreasing the number of documents. No further preprocessing of documents was performed, as the analysis was to utilize transformer models, which leverage both lexical and non-lexical linguistic information. This process resulted in a corpus of 499.390 documents of statements from Danish politicians which were to be analyzed.

Topic Modeling {MSRP}

In order to answer the research questions topic modeling was done using BERTopic (Grootendorst, 2022). Documents were embedded using two pre-trained sentence-transformer models; paraphrase-multilingual-MiniLM-L12-v2, BERTopics built-in multilingual model optimized for danish (Grootendorst, 2024), and Munin 7B Alpha, a model finetuned from Mistral-7B-v0.1 using the Danish Gigaword Corpus (Danish Foundation Models Team, 2024). Model performance was assessed by the authors, and it was deemed that BERTopics built-in multilingual model created the most relevant and coherent topic clusters, in line with findings from Navarretta & Hansen (2023). As such, the following analysis will concern results from this model.

The model clustered statements into 3624 topics. The 343 most frequent topics were further manually clustered into 20 topic clusters[3] taken from D. H. Hansen et al. (2019), which are displayed in Appendix 1. The cut-off of 343 topics was chosen as these encompassed more than 80% of statements and further manual clustering had diminishing returns. Clustering was done independently by the authors, by inspecting the representative document and four words for each topic, after which clusters were compared and discrepancies were resolved. This process may have introduced biases in the analysis and will be further discussed in the limitations section.

During manually clustering of the topics, 168 of the original 343 most frequent topics were deemed irrelevant, as they were clusterings of stopwords and names. This resulted in the 20 topics covering a corpus of 126021 statements. The distribution of statements over time and topics can be seen in the appendix (*Appendix 2 and Appendix 3)*.

The statements were merged with demographic information, where the speaker's age was computed at the time of each statement. After preprocessing demographics were distributed as; age at statement (*M = 48 years, SD = 11,25)*, gender (63% male, 37% female) and party (*see Appendix 4)*.

Sentiment Analysis {FRFL}

To explain the tone of debates, sentiment analysis of document polarity was employed. Polarity was classified as either positive, neutral or negative. This was carried out with the Python package transformers (Wolf et al., 2020) using da-sent-xlm; a fine-tuned version of the sentiment classification model XLM-RoBERTa optimized for Danish (Snæbjarnarson, 2024). This model has a tensor-limit of 512 meaning documents exceeding this amount of characters were truncated. The resulting distribution of sentiment polarity can be found in Appendix 5.

---

[2] Specifically: “Ministeren.” (*“The minister.”), “Spørgeren.” (“The asking.”)* and *“Formanden.” (“The Speaker.”)*

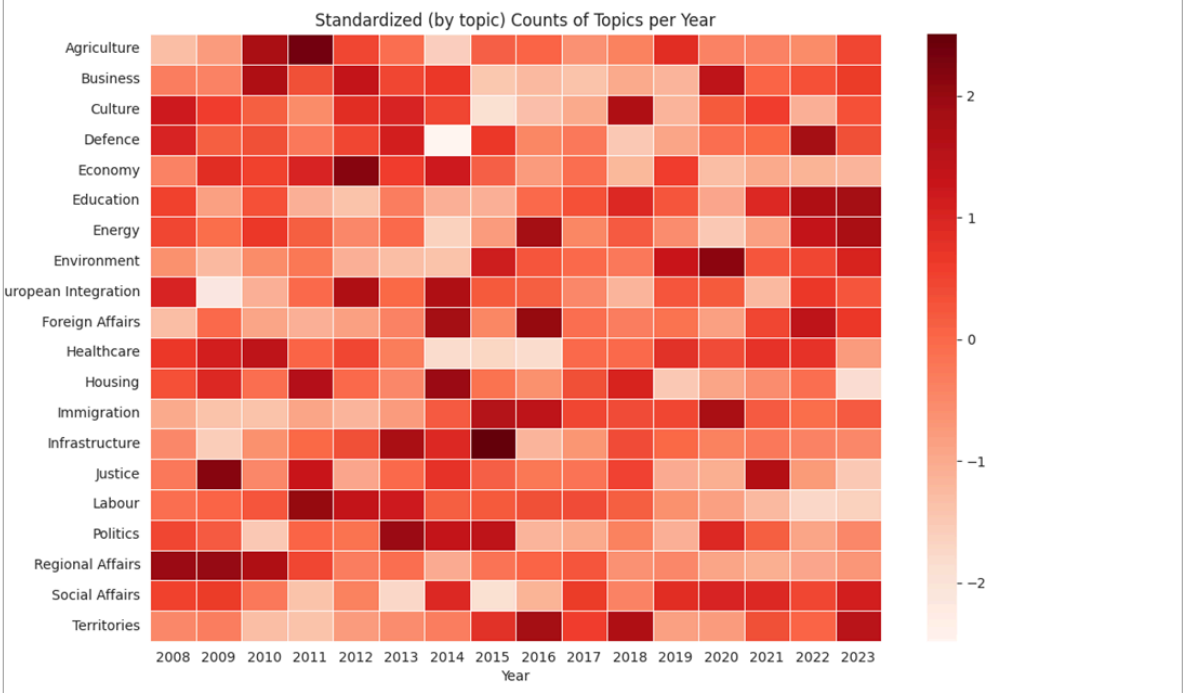
[3] Note that two clusters of the 22 topic clusters from D. H. Hansen et al. (2019) were not used, as “Government Operations” and “Politics” were joined in the current analysis, and no topics fell under “Personal Rights”.

Results

1. Main Trends of Danish Parliamentary Debates {MSRP}

To investigate how topics in parliamentary debates have evolved over time a heatmap (*Figure 1*) was created, which has been z-scored first by year and then by topic, to account for the different overall counts of statements for each individual year and topic, ultimately showing how much a given topic was part of one year's discourse in relation to previous years.

Figure 1: Evolution of Topics



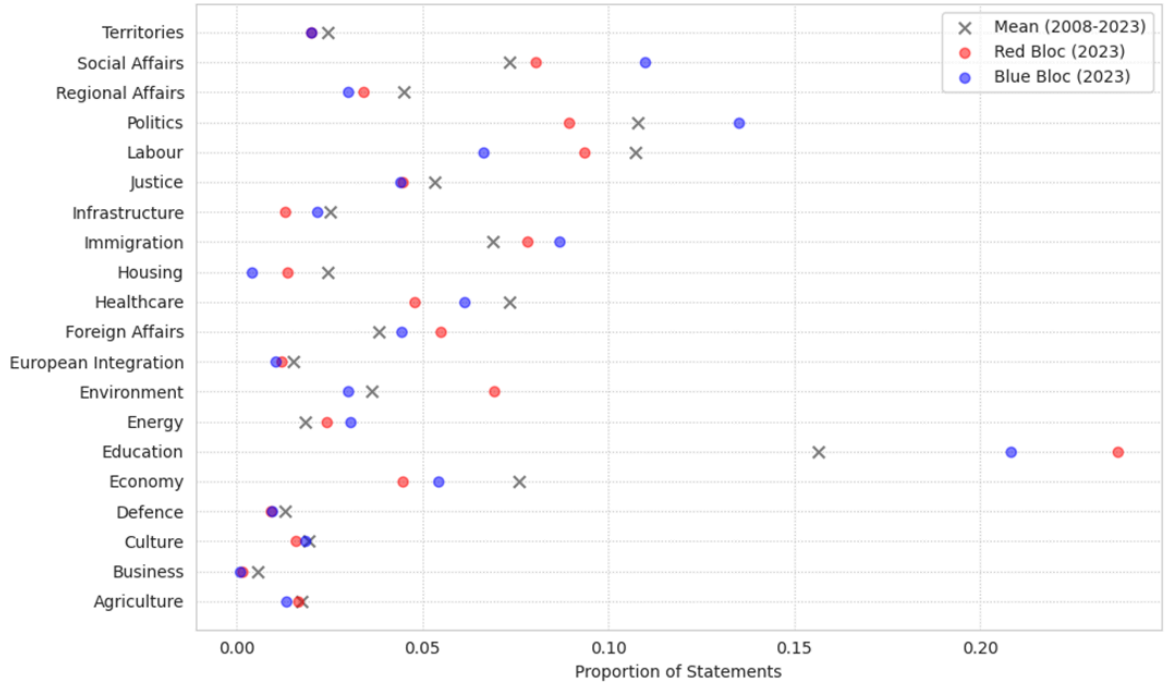
As a general tendency, it seems politicians speak less about economy, labour, agriculture and regional affairs, and more about education, energy, environment, foreign affairs, immigration, and social affairs, in line with previous studies (Green-Pedersen, 2013; Green-Pedersen & Kosiara-Pedersen, 2020; Skjæveland, 2005).

The heatmap also reflects specific political and societal events. An obvious example is the Covid-19 outbreak and subsequent lockdown, which align with higher scores in the topic "healthcare" from 2019-2022. Covid-19 can also be seen affecting "business" with an increase in discussions reflecting that many businesses were temporarily shut down.
Other examples include an increase in the topics "defence", "foreign affairs", and "EU integration" from 2015 to 2017, aligning with the refugee crisis in 2015, terrorist attacks in Paris, Brussels, and Denmark, as well as wars in the Middle East. "Defence" is again heavily discussed in 2022 coinciding with Russia's invasion of Ukraine in early 2022.
Further, the topic "environment" has an increase in 2020, aligning with the election of 2019 being referred to as the "climate election" (K. M. Hansen & Stubager, 2021). The topic "energy" has an increase in 2016 and 2023, aligning with the collapse of oil prices in 2015 (Stocker et al., 2018), and

the energy crisis in 2022 (Gilbert et al., 2021). Looking at these patterns, it seems like there is a delay in when the topics are discussed in the parliament compared to when the event happens, reflecting findings in the previously mentioned study by Bourgeois et al in 2022.

Topics Distribution by Bloc {FRFL}

Figure 2: Topics Distribution by Bloc



To further explicate the development of topics, Figure 2 illustrates the proportion of statements between red and blue bloc distributed across topics in 2023 compared to the mean across the entire time frame. The topic “environment” is noticeable, as it is clear red bloc speaks more about the environment than blue bloc, aligning with findings from Navarretta & Hansen (2023).

Environment Statements by Parties {MSRP}

Figure 3: Environment Statements by Parties

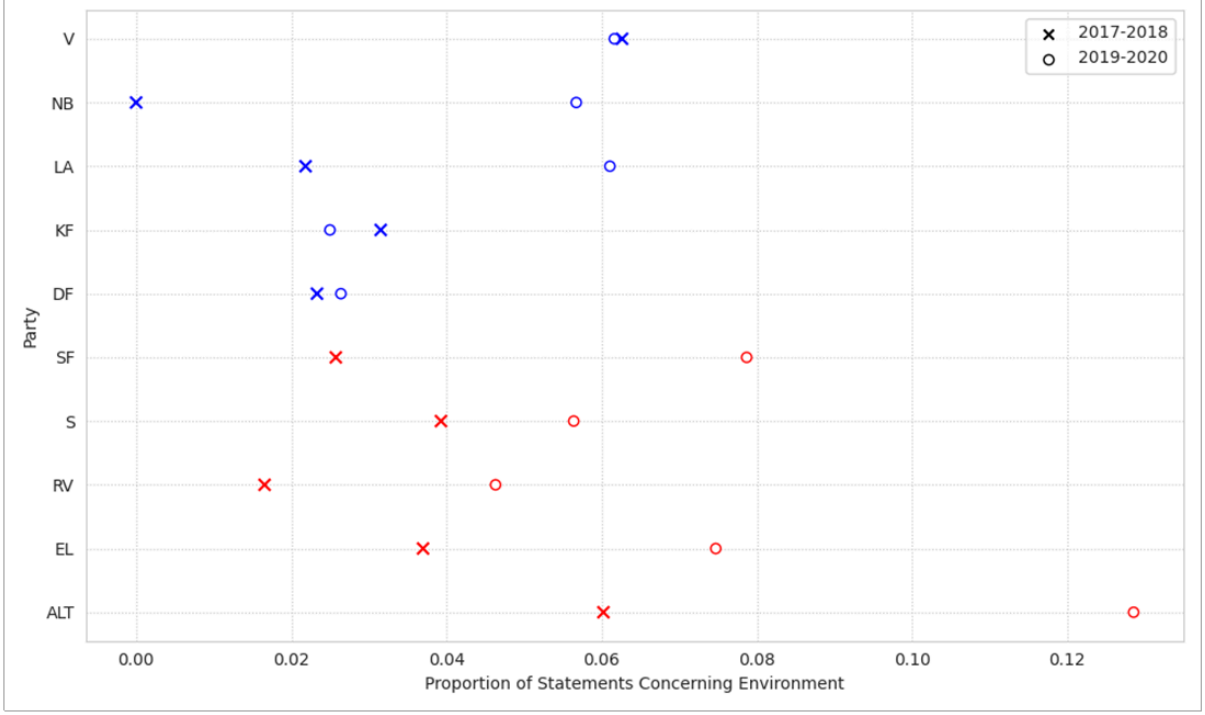
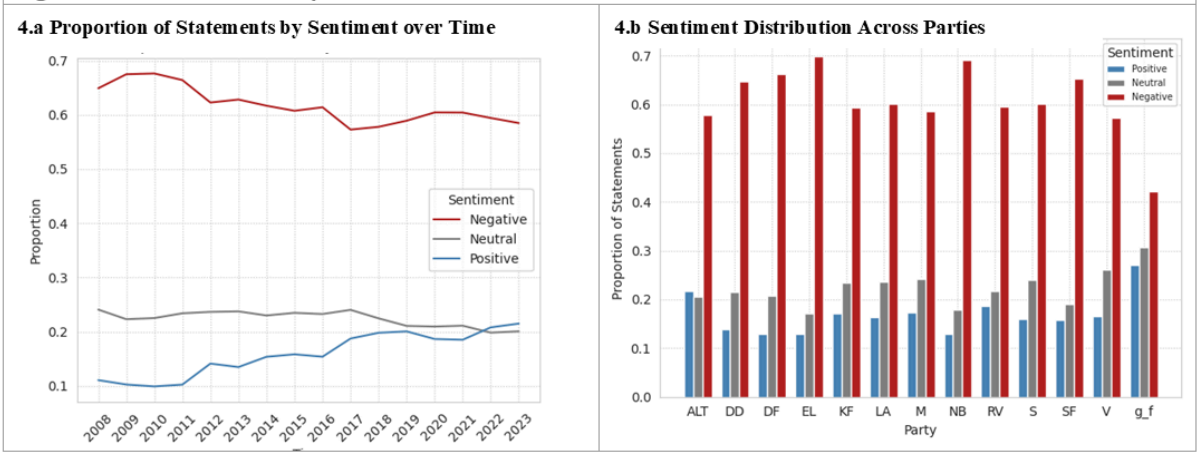


Figure 1 showed an increase in statements concerning "environment" in 2019 and 2020 coinciding with the climate election in November 2019. Additionally, Figure 2, showed that red bloc spoke more about the environment compared to blue bloc. Therefore, an elucidation of the topic environment was made looking into how much parties discussed the environment in the period from 2017 to 2020, illustrated in Figure 3. The crosses represent the proportions of statements concerning the environment relative to parties' overall statement count in 2023 before the climate election (2017-2018), and the circles represent the proportions after the climate election (2019-2020). For translation of the party abbreviation, refer to Appendix 6.

Figure 3 indicates that all parties, except Konservative Folkeparti (*Conservative People's Party*), have spoken more about the environment following the climate election. The tendency displayed in Figure 2, that red bloc has spoken more about climate than blue bloc, seems to be affected by the climate election, where red bloc exhibits a greater disparity in how much they have spoken about the environment before and after the election. As a final note, Alternativet seems to speak more about the environment compared to other parties, aligning with them being founded as a response to the climate crisis as outlined by Green-Pedersen & Kosiara-Pedersen (2020).

The Tone of the Parliamentary Debate {FRFL}

Figure 4: Sentiment Analysis



To further explore tendencies in parliamentary debates, sentiment analysis was used to investigate the tone of the debates. Figure 4.a visualizes the proportion of statements by sentiment over from 2008 to 2023, with a line for each sentiment polarity. The plot suggest that the tone of parliamentary debates is predominantly negative. However, there seems to be a trend of debates steadily becoming increasingly positive from 2008 when 11,04% of the statements were positive to 21,47% in 2023. This mimics results from sentiment analysis on parliamentary debates from the British Parliament from 1909-2013, which found the positivity of statements' sentiment to be increasing (Rheault et al., 2016), without drawing any conclusions as to why. Similarly, more research would be required to explain the trend seen in Figure 4.a.

Sentiment proportions across parties can be seen in Figure 4.b. Apart from Alternativet and politicians from Greenland and the Faroe Islands sentiment proportions appear mostly even. These outliers can be explained in Alternativets case by their political manifesto, in which they want to serve as a positive counterpoint in the political debate (*Manifest*, 2024). The deviation of Greenlandic and Faroese politicians' sentiment distribution may be due to them being minorities. This can be explained by the theory "Doing Difference" by West & Fenstermarker (1995), which serves as an intersectional theory explaining how minorities are always held accountable for their actions, hence having a limited latitude for their actions. The discrepancy can therefore be due to them overcompensating, making it particularly challenging for them to express negative emotions, as they try to avoid facing stigmatization.
Though distributions seem even, Figure 4.b highlights a distinction between parties that have held government positions (KF, LA, M, RV, S, V, SF) and those that have not (DD, EL, DF, NB, ALT) between 2008-2024, with parties consistently outside the government having more negative statements. A point which can be elaborated upon when investigating differences in government and opposition.

2. Government and Opposition {MSRP}

Figure 5: Government vs. Opposition

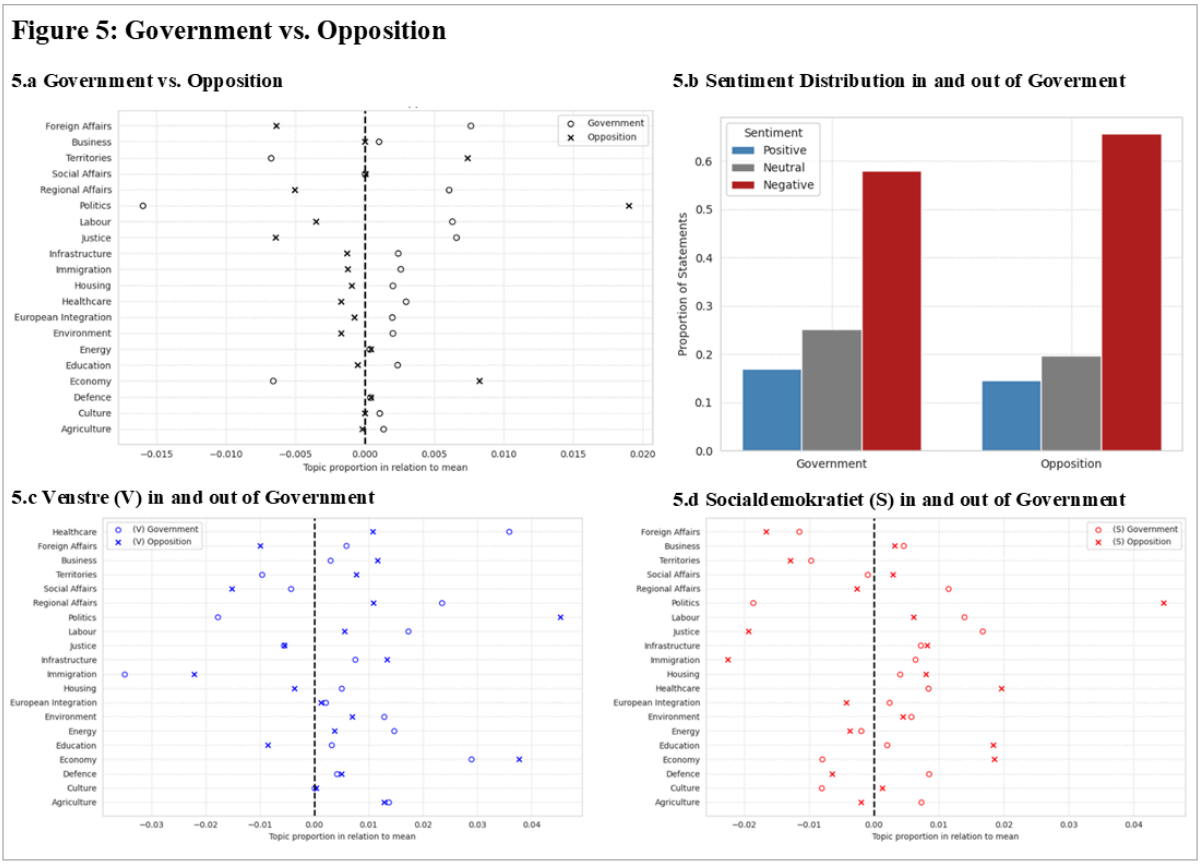


Figure 5.a illustrates variances in statements between government and opposition parties across topics between 2008 and 2024. The topic proportion is zero-centered, where the dotted line represents the mean. Notably, topics “politics” and “economy” are discussed more by parties in the opposition. Helms (2008, page 9) defines the opposition's role as consisting of three tasks; criticizing the government, scrutinizing and checking governmental actions and policies, and representing a credible “alternative government”. This could explain the heightened debates of these topics, as the opposition parties are critical regarding the government and its internal politics, as their role is to hold the government parties responsible. Furthermore, regarding “economy”, Socialdemokratiet and Venstre are historically separated in their position concerning redistribution politics (Slothuus, 2003), which could explain the heightened discussion of “economy” from the opposition.
The critique from the opposition can also benefit from being examined through the use of sentiment analysis, where the prevalence of criticism from the opposition is indicated by a higher proportion of negative statements compared to parties in government. This is depicted in Figure 5.b, showing the proportion of positive (blue), neutral (grey), and negative statements (red) between 2008-2024, indicating that parties in the opposition exhibit higher levels of negativity compared to parties in the government. These findings align with previous studies using sentiment analysis comparing European opposition and government parties in parliamentary debates, which also found more negative statements from opposition parties (Proksch et al., 2019; Rudkowsky et al., 2018).

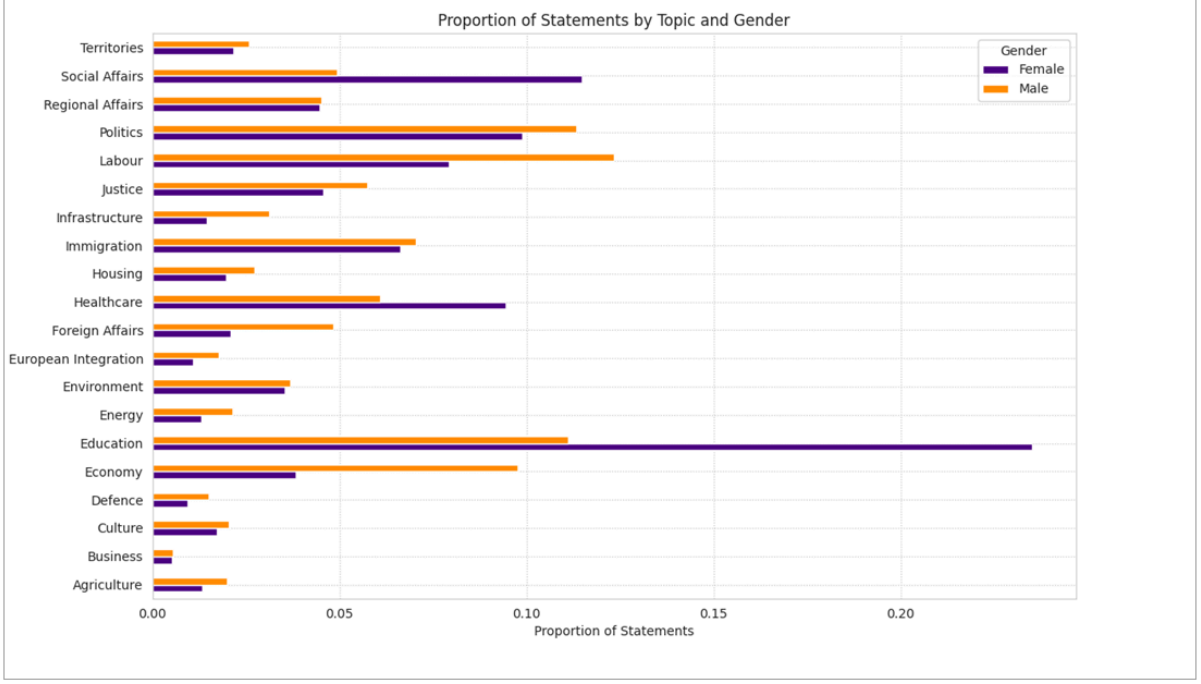
Further investigation into the disparities between government and opposition explored the differences in when the leading parties of red and blue bloc, respectively Socialdemokratiet and Venstre, were in and out of government, as illustrated in Figure 5.c and 5.d. The topic proportion is again zero-centered. The figures appear to display an alignment between the topics spoken more about when

in opposition and parties' key issues. Socialdemokratiet speaks more about social affairs, healthcare, education, and culture when being in the opposition aligning with some of their key issues; equality and wealthfare (*Politik | Socialdemokratiets politik og mærkesager*, 2024), whereas Venstre speaks more about business, economy and infrastructure when being in the opposition aligning with some of their key issues; financial liberty and immigration (*Mere frihed. Flere muligheder*, 2024). These findings are in line with previous studies investigating the dynamic between government and opposition, in which it is argued that opposition parties are freer to focus on issues that resonate with their voters, whereas the government parties are responsible for creating solutions to a variety of societal issues (Green-Pedersen, 2013; Vliegenthart & Walgrave, 2011). However, it would be necessary to investigate this further in order to draw stronger conclusions as the effect size appears small.

3. The Influence of Gender and Age on Topics {FRFL}

The Influence of Gender on Topics

Figure 6: Proportion of Satements by Topic and Gender

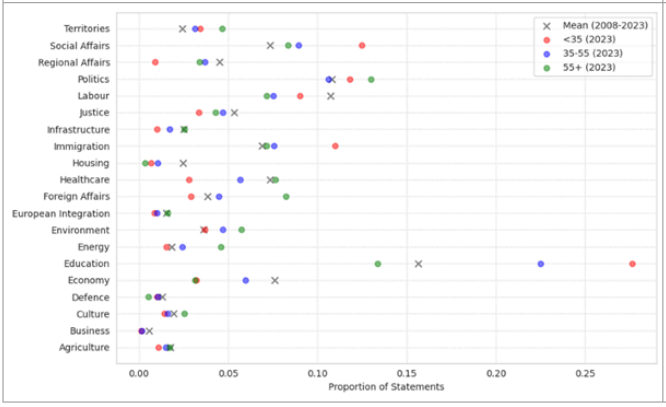


To explore gender differences, a bar plot was created showing the proportion of topics discussed by respectively men and women (*Figure 6)*. The purple bar represents statements from female politicians, and the orange bar represents statements from male politicians. These results suggest women are more likely to discuss topics related to "social affairs", "healthcare", and "education", whereas men are more likely to discuss topics related to "labour", "foreign affairs", and "economy". These findings align with previous studies of the British Parliment and voters (Blaxill & Beelen, 2016; K. M. Hansen, 2019), indicating that politicians' party affiliations are influenced by gender.

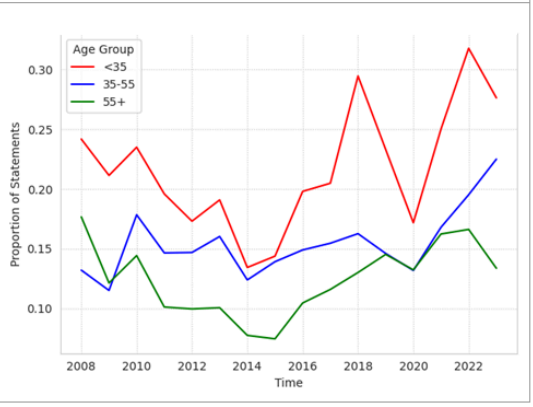
The Influence of Age on Topics

Figure 7: Influence of Age

7.a Influence of Age on Topics



7.b Proportion of Statements about Education by Age



Concerning the demographic, age, Figure 7.a illustrates the proportion of statements across topics in 2023 in three different age groups; under 35, between 35-55, and above 55. Age groups were taken from K. M. Hansen (2019)[4] research regarding voters, however, these were slightly changed due to the young cohorts and elderly people being underrepresented in the parliament. The young age group is defined as politicians under 35 years, and the elderly age group is defined as politicians above 55 years.

Looking at the plot, it seems like there is a general tendency of younger politicians being more concerned with education compared to elders, and that the older the politicians are the less they speak about education. This is investigated further in Figure 7.b illustrating the proportion of statements concerning education between the three age groups from 2008 to 2023. It shows young politicians consistently valuing education higher the last 14 years. These findings align with previous research by Metz (2002) and Stockemer & Sundström (2018).

Discussion

Discussion of Results {MSRP}

The preceding analysis aimed to assess how topic modeling and sentiment analysis can be used as tools in exploring Danish parliamentary debates. The following section will summarize the results and discuss natural language processing's applicability.

Figure 1 illustrates how topics have evolved from 2008 to 2023 reflecting several events such as Covid-19, the climate election, energy crises, refugee crisis, and the general shift in Danish politics moving focus from economy- and finance-related topics to more social- and environmental-related topics. These results seem to validate topic modeling as an approach for successfully extracting topics from Danish Parliamentary debates. However, it was found that when interpreting topics by year, the effects of political and societal events can often first be seen one year after the onset of an event such as the climate election or refugee crisis, as parliamentary debates seem to follow their own cycle.

[4] Which categorized age as; under 30, between 30-60, and above 60.

From Figure 1 it seemed that the topic environment was discussed more in recent years. When investigating these trends further, in Figure 2, it was found that in 2023 this seemed to be an effect caused by red bloc. When this was explored further, by plotting how the proportion of statements regarding the environment before and after the climate election (*Figure 3)*, it seemed like this trend may have been a cause of the climate election, as red bloc in the years following exhibit greater disparity in their proportions of statements about the environment compared to blue bloc. This further illustrates how topic modeling can be suitable for detecting various interacting tendencies on both macro- and micro-level.
When investigating sentiment, parliamentary debates were found to be increasingly positive, yet there seems no apparent explanation for this development, and would as such require further research. This serves as an example of sentiment analysis being a useful tool in discovering underlying tendencies in politics, revealing new questions to be researched.

Investigation of other factors relevant to political stance, namely political position, age and gender further exemplifies NLP's applicability. It was found that opposition parties speak more about internal politics and economy, by which they criticize and thereby hold the government responsible for their politics, which is supported by sentiment analysis showing that the opposition parties are more negative compared to parties in the government. This elucidates how topic modeling and sentiment analysis can be used in interplay, resulting in more nuanced interpretations of findings. Further, the results of gender and age indicate an influence on the topics discussed by politicians, in which women tend to speak more about welfare-related topics, men more about finance, and young cohorts more about education. These findings, especially the large differences found when examining gender, warrant further research and again display natural language processing's ability to expose general tendencies inspiring further investigation.

In answering the main research question of this paper, the various findings suggest that the use of topic modeling and sentiment analysis are suitable methods for exploring tendencies in Danish parliamentary debates. It is able to detect tendencies at various levels; time, gender, age, topic, bloc and party - creating clarity in the trends of the political system. The aforementioned personalization trend to a greater extend entails politicians focusing on how they present themselves on social media as opposed to their parliamentary work. This being the case, the ability to ascertain the actual work being done is more valuable than ever. However, it is worth mentioning that while the parliamentary debates reflect what is happening inside Christiansborg, the debates are broadcast. Politicians are undoubtedly aware of this, which might influence how they present themselves, compared to how they present themselves behind closed doors, for example in internal meetings.

Limitations and Future Directions {FRFL}

The topic modeling and sentiment analysis processes did not come without their limitations. When classifying topics, BERTopic only assigned one topic to each statement, despite longer statements potentially containing multiple topics. This could have been accommodated by splitting longer statements, like speeches, into smaller segments.
Another consideration is that the model used in the current analysis was not explicitly trained on Danish political texts. Here a possibility would be to do a model comparison between BERTopics multilingual embedding model and a model trained on the annotated corpus from D. H. Hansen et al. (2019), which was created for training models for classification of Danish Parliamentary debates. This would be an interesting prospect, which may produce more relevant topics and topic counts.

In terms of manual topic clustering, this was done by the authors, and although done independently, might still have introduced biases. To accommodate this, it would be sensible to do inter-coder reliability in future studies, to ensure a more objective classification.

Turning to sentiment analysis, each statement was simply assigned a polarity label by the model. However, each statement was also assigned a sentiment score, which reflects how the level of confidence in the assigned polarity - this was not taken into account, and the categorization might be affected as a result. A possibility would have been to create a threshold for assigning the value, ensuring a level of confidence in polarity labels were correctly assigned. The limitation of using this method is that not all statements would be taken into account in analysis. A comparison between these approaches could have been beneficial.
A further consideration concerning sentiment analysis is that only one value was assigned to each statement. As such shifts in sentiments during statements were not detected. Here sentiment analysis could, like topic modeling, have benefited from splitting longer statements.

A paper investigating sentiment analysis in parliamentary debates argues that sentiment models have difficulties assigning sentiment values in parliamentary speeches which are not very consistent in their sentiment expressions. Thereby concluding that the sentiment models without modifying them to parliamentary debates is not very sufficient (Meden, 2022). This calls for further investigation of the results found using sentiment analysis in order to draw stronger conclusions on the current findings.

An additional factor to consider is that party changes were not taken into account, as each parliament member was assigned only to their latest party. Members without party affiliation were assigned to the party they served the longest between 2007 and 2024, as most of their statements were made during that period for the specific party. This especially applies to members of the party, Nye Borgerlige, which was disbanded in 2024. Members of this party who became independent were assigned to Nye Borgerlige, despite its dissolution. Members who changed party, were assigned their current party. Given the increase in party change in recent years, it could be beneficial to take party change into account for further investigations.

Data Availability {MSRP}

As this paper serves as an exploratory paper, aiming to inspire researchers from various fields in politics and computer science to expand on our investigation. Most of the findings in this paper could be further supported by a more comprehensive analysis, in which it could be beneficial to incorporate statistical tests and additional underlying explanations from political science. As such, data will be available at GitHub (https://github.com/magnusseverinrp/danish-parlimentary-debates/tree/main).

Conclusion {FRFL & MSRP}

The aim of this paper was to investigate how topic modeling and sentiment analysis can be used as tools in exploring tendencies in Danish parliamentary debates. In doing so, exploring how topics have evolved, how they are distributed between blocs, and how single events, such as the 2019 climate election can affect these trends. Furthermore, how being in government or opposition affects the topics and tone, and lastly the influence of gender and age on the topics were discussed. Topic modeling was found capable of reflecting several societal events, validating topic modeling as an approach to exploring Danish parliamentary debates. It can be used to reveal tendencies at various levels; time, gender, topic, bloc and party levels, and by combining these, it is possible to extract how

these interact. Sentiment analysis was likewise found to be a useful tool in exploring general tendencies of the tone in the parliamentary debates, finding an increase in positive statements. Furthermore, it can be used in interplay with topic modeling, creating more comprehensive findings. The findings of this paper serve as an exploration of topic modeling's applicability and have aimed to inspire researchers from other fields to expand on our investigation, thereby creating more transparency in the political system.

Bibliography

Abercrombie, G., & Batista-Navarro, R. (2020). Sentiment and position-taking analysis of parliamentary debates: A systematic literature review. *Journal of Computational Social Science*, *3*(1), 245–270. https://doi.org/10.1007/s42001-019-00060-w

Abrami, G., Bagci, M., Hammerla, L., & Mehler, A. (2022). *German Parliamentary Corpus (GerParCor)* (arXiv:2204.10422). arXiv. http://arxiv.org/abs/2204.10422

Blaxill, L., & Beelen, K. (2016). Women in Parliament since 1945: Have they changed the debate? *History & Policy*. https://www.historyandpolicy.org/index.php/policy-papers/papers/women-in-parliament-since-1945-have-they-changed-the-debate

Bourgeois, N., Pellet, A., & Puren, M. (2022). *Using Topic Generation Model to explore the French Parliamentary Debates during the early Third Republic (1881-1899)*. https://ceur-ws.org/Vol-3133/paper03.pdf

Budge, I., Klingemann, H.-D., Volkens, A., Bara, J., Tanenbaum, E., Fording, R., Hearl, D., Kim, H., Mcdonald, M., & Mendes, S. (2023). *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments 1945—1998*. https://doi.org/10.1093/oso/9780199244003.001.0001

Contreras, K., Verbel, G., Sanchez, J., & Sanchez-Galan, J. E. (2022). Using Topic Modelling for Analyzing Panamanian Parliamentary Proceedings with Neural and Statistical Methods. *2022 IEEE 40th Central America and Panama Convention (CONCAPAN)*, 1–6. https://doi.org/10.1109/CONCAPAN48024.2022.9997766

Dahal, B., Kumar, S. A. P., & Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. *Social Network Analysis and Mining*, *9*(1), 24. https://doi.org/10.1007/s13278-019-0568-8

Danish Foundation Models Team. (2024, april 11). *Danish-foundation-models/munin-7b-alpha · Hugging Face*. https://huggingface.co/danish-foundation-models/munin-7b-alpha

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.886498

Enli, G. (2015). *Mediated Authenticity: How the Media Constructs Reality*. https://doi.org/10.3726/978-1-4539-1458-8

Enli, G., & Moe, H. (2013). INTRODUCTION TO SPECIAL ISSUE. *Information*, *16*. https://doi.org/10.1080/1369118X.2013.784795

Enli, G. S., & Skogerbø, E. (2013). Personalized Campaigns in Party-Centred Politics: Twitter and Facebook as arenas for political communication. *Information, Communication & Society*, *16*(5), 757–774. https://doi.org/10.1080/1369118X.2013.782330

Gilbert, A., Bazilian, M. D., & Gross, S. (2021). *The emerging global natural gas market and the energy crisis of 2021-2022*. https://www.brookings.edu/wp-content/uploads/2021/12/FP_20211214_global_energy_crisis_gilbert_bazilian_gross.pdf

Green-Pedersen, C. (2013). *Partier i nye tider. Den politiske dagsordensfastsættelse i Danmark*. *45. årg. nr. 1*, 6–17.

Green-Pedersen, C., & Kosiara-Pedersen, K. (2020). Det danske partisystem mellem kontinuitet og forandring. I *De danske politiske partier* (s. 215–234). Tidsskriftet Politica.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*.

Grootendorst, M. (2024). *BERTopic—BERTopic*. https://maartengr.github.io/BERTopic/api/bertopic.html

Hansen, D. H., Navarretta, C., Offersgaard, L., & Wedekind, J. (2019). Towards the Automatic Classification of Speech Subjects in the Danish Parliament Corpus. *CEUR Workshop Proceedings, 2364*, 166–174.

Hansen, K. M. (2019). *Sådan fordeler partiernes vælgere sig på køn, alder og uddannelse*. https://www.altinget.dk/aeldre/artikel/se-tallene-saadan-fordeler-partiernes-vaelgere-sig-paa-k

oen-alder-og-uddannelse

Hansen, K. M., & Stubager, R. (2017). *Oprør fra udkanten: Folketingsvalget 2015*. Djøf Forlag.

Hansen, K. M., & Stubager, R. (Red.). (2021). *Klimavalget: Folketingsvalget 2019* (1. udgave, 1. oplag). Djøf Forlag.

Hansen, K. M., & Stubager, R. (2024). *Partiledernes kamp om midten: Folketingsvalget 2022*. Djøf Forlag; Studier i dansk politik.

Helms, L. (2008). Studying Parliamentary Opposition in Old and New Democracies: Issues and Perspectives. *The Journal of Legislative Studies*, *14*, 6–19. https://doi.org/10.1080/13572330801920788

Hyvönen, E. (2022). *Plenary Speeches of the Parliament of Finland as Linked Open Data and Data Services*.

Ivanusch, C., Burst, T., & Zehnter, L. (2022). *Challenges and opportunities for the comparative study of political text types: Testing computer-based topic classification approaches across political party texts*. OPTED.

Kozbagarov, O., Mussabayev, R., & Mladenovic, N. (2021). A New Sentence-Based Interpretative Topic Modeling and Automatic Topic Labeling. *Symmetry*, *13*(5), Article 5. https://doi.org/10.3390/sym13050837

Lazreg, L. A. (2023). *Topic Modelling for Metadata Extraction and Generation from Norwegian Parliamentary Texts* [Master thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3097127

LI, X., KAWAMURA, K., & KIMURA, Y. (2024). Quantifying the Activities of Local Assembly Members in Japan: Recent Advances and an Approach Using the BERTopic Model. *Interdisciplinary Information Sciences*, *30*. https://doi.org/10.4036/iis.2024.R.03

Loftis, M., & Mortensen, P. B. (2020). Collaborating with the Machines: A hybrid method for classifying policy documents. *Policy Studies Journal*, *48*(1), 184–206. https://doi.org/10.1111/psj.12245

Louis, A. (2020, juli 7). A Brief History of Natural Language Processing—Part 2. *Medium*. https://medium.com/@antoine.louis/a-brief-history-of-natural-language-processing-part-2-f5e

575e8e37

Mandatfordelingen. (2024). Folketinget. https://www.ft.dk/da/medlemmer/mandatfordelingen

Manifest. (2024). Alternativet. https://alternativet.dk/politik/manifest

Meden, K. (2022). Speech-level Sentiment Analysis of Parliamentary Debates using Lexicon-based Approaches. *Digital Humanities*.

Mere frihed. Flere muligheder. (2024). Venstre - Mere frihed. Flere muligheder. https://www.venstre.dk/politik/maerkesager

Metz, D. (2002). The Politics of Population Ageing. *The Political Quarterly*, *73*(3), 321–327. https://doi.org/10.1111/1467-923X.00472

Navarretta, C., & Haltrup Hansen, D. (2020). Identifying Parties in Manifestos and Parliament Speeches. I D. Fišer, M. Eskevich, & F. de Jong (Red.), *Proceedings of the Second ParlaCLARIN Workshop* (s. 51–57). European Language Resources Association. https://aclanthology.org/2020.parlaclarin-1.10

Navarretta, C., & Hansen, D. H. (2023). According to BERTopic, what do Danish Parties Debate on when they Address Energy and Environment? I C. Klamm, G. Lapesa, V. Gold, T. Gessler, & S. P. Ponzetto (Red.), *Proceedings of the 3rd Workshop on Computational Linguistics for the Political and Social Sciences* (s. 59–68). Association for Computational Lingustics. https://aclanthology.org/2023.cpss-1.6

Nørtoft, A. M. (2019). *Næsten halvdelen af folketingsmedlemmerne har en lang videregående uddannelse*. https://www.dst.dk/da/Statistik/nyheder-analyser-publ/bagtal/2019/2019-09-27-naesten-halvdelen-af-folketingsmedlemmerne-har-en-lang-videregaaende-uddannelse

Oda.ft.dk. (2024). https://oda.ft.dk/

Perak, B., & Rodik, F. (2018). Building a corpus of the Croatian parliamentary debates using UDPipe open source NLP tools and Neo4j graph database for creation of social ontology model, text classification and extraction of semantic information. *Digital Humanities*.

Politik | Socialdemokratiets politik og mærkesager. (2024). https://www.socialdemokratiet.dk/det-vil-vi/

Proksch, S.-O., Lowe, W., Wäckerle, J., & Soroka, S. (2019). Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches. *Legislative Studies Quarterly*, *44*(1), 97–131. https://doi.org/10.1111/lsq.12218

Referater. (2016, november 30). Folketinget. https://www.ft.dk/da/dokumenter/dokumentlister/referater

Rheault, L., Beelen, K., Cochrane, C., & Hirst, G. (2016). Measuring Emotion in Parliamentary Debates with Automated Textual Analysis. *PLOS ONE*, *11*(12), e0168843. https://doi.org/10.1371/journal.pone.0168843

Richardson, L. (2007). *Beautiful soup documentation* [Software].

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., & Sedlmair, M. (2018). More than Bags of Words: Sentiment Analysis with Word Embeddings. *Communication Methods and Measures*, *12*(2–3), 140–157. https://doi.org/10.1080/19312458.2018.1455817

Severin-Nielsen, M. K. (2024). Electoral campaigning in a hybrid media environment: A case study of two Danish party leaders' social media campaigns and online news presence during the 2022 parliamentary elections. *Journalistica*, *18*(1), Article 1. https://doi.org/10.7146/journalistica.v18i1.138048

Skidmore-Hess, D. (2003). The Danish Party System and the Rise of the Right in the 2001 Parliamentary Election. *International Social Science Review*, *78*(3/4), 89–110.

Skjæveland, A. (2005). Dimensionaliteten i Folketinget. *Politica*, *37*(4), Article 4. https://doi.org/10.7146/politica.v37i4.69855

Skogerbø, E., & Larsson, A. O. (2021). Comparing Twitter and Instagram as platforms for party leader communication: Findings from the 2017 Norwegian election. *First Monday*. https://doi.org/10.5210/fm.v26i10.10638

Slothuus, R. (2003). Partier med omtanke: Socialdemokratiets og Venstres vælgerstrategier. *Politica*, *35*(2), 78–94.

Snæbjarnarson, V. (2024). *Vesteinn/danish_sentiment · Hugging Face*. https://huggingface.co/vesteinn/danish_sentiment

Stockemer, D., & Sundström, A. (2018). Age representation in parliaments: Can institutions pave the

way for the young? *European Political Science Review*, *10*(3), 467–490. https://doi.org/10.1017/S1755773918000048

Stocker, M., Baffes, J., Some, Y. M., Vorisek, D., & Wheeler, C. M. (2018). *The 2014–16 Oil Price Collapse in Retrospect: Sources and Implications*. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-8419

Taboada, M. (2016). Sentiment Analysis: An Overview from Linguistics. *Annual Review of Linguistics*, *2*(Volume 2, 2016), 325–347. https://doi.org/10.1146/annurev-linguistics-011415-040518

Toft, J. (2022, august 12). Næsten hver femte politiker har forladt sit parti i Folketinget. *Vid&Sans*. https://vidogsans.dk/naesten-hver-femte-politiker-har-forladt-sit-parti-i-folketinget/

Tucker, J., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3144139

Umamaheswaran, S., Dar, V., Sharma, E., & Kurian, J. S. (2023). Mapping Climate Themes From 2008-2021—An Analysis of Business News Using Topic Models. *IEEE Access*, *11*, 26554–26565. https://doi.org/10.1109/ACCESS.2023.3256530

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vliegenthart, R., & Walgrave, S. (2011). Content Matters: The Dynamics of Parliamentary Questioning in Belgium and Denmark. *Comparative Political Studies*, *44*(8), 1031–1059. https://doi.org/10.1177/0010414011405168

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

WEST, C., & FENSTERMAKER, S. (1995). DOING DIFFERENCE. *Gender & Society*, *9*(1), 8–37. https://doi.org/10.1177/089124395009001002

Wolf, T., Debut, L., Sanh, V., Chaumond, J., & Delangue, C. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

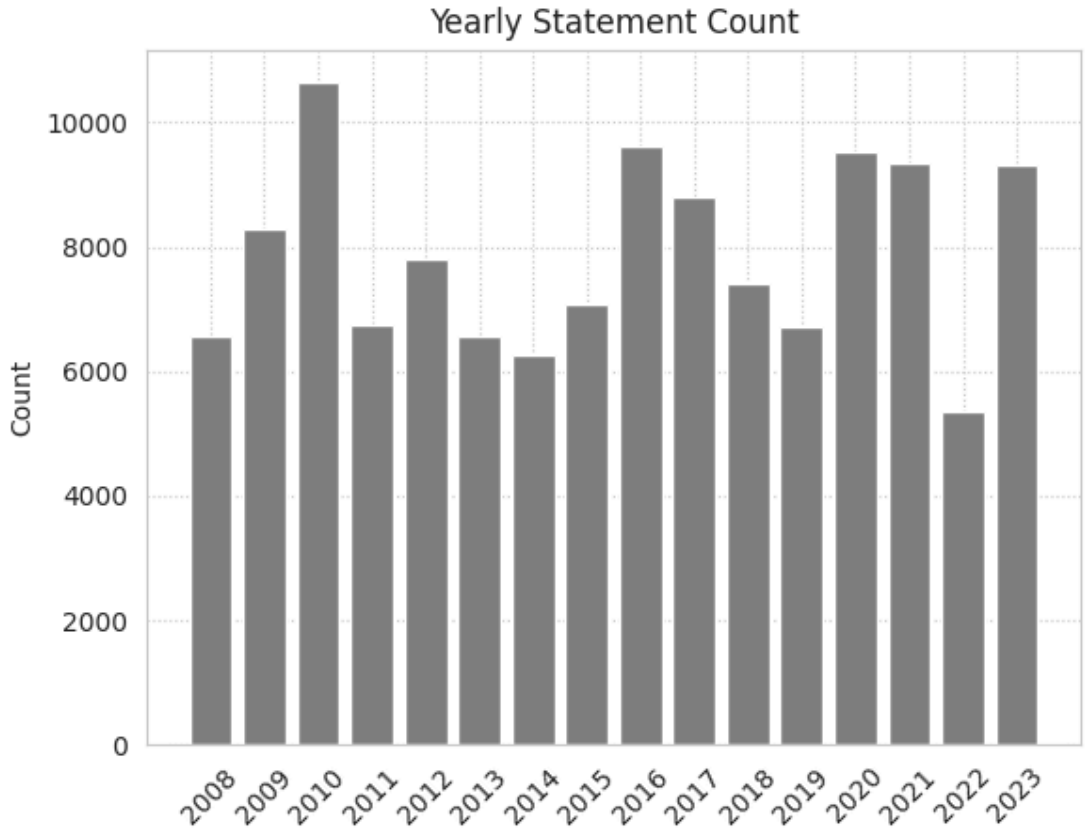
https://github.com/huggingface/transformers

Appendix

Appendix 1: Topics and Representative Words

Topic	Representative words	English translation
Agriculture	landbrug_landbruget_fiskeri_landmænd	*agriculture, fishery, farmers*
Business	turisme_turismeerhvervet_reklamer_forbrugeren	*tourism, tourism industry, advertisment, consumers*
Culture	sprog_radio_fyrværkeri_folkekirken	*language, radio, fireworks, the nation church*
Defence	kampfly_afghanistan_taleban_fly	*fighter jets, Afghanistan, Taliban, jets*
Economy	skattelettelser_skattereform_skat_skatten	*tax cuts, tax reform, taxes, tax*
Education	elever_folkeskolen_uddannelse_studerende	*pupils, elementary school, education, students*
Energy	energi_vedvarende_vindmøller_atomkraft	*energy, sustainable, wind turbines, nuclear power*
Environment	vandløb_elbiler_klimarådet_nox	*watercourses, electric cars, climate counsil, nitrogen oxide*
European Integration	europol_eu_europa_europæiske	*european politics, eu, europa, european*
Foreign Affairs	israel_tyrkiet_nordiske_hamas	*Israel, Turkey, nordic, Hamas*
Healthcare	patienter_sundhedsvæsen_læger_patienterne	*patients, healtcare, doctors*
Housing	boliger_almene_boligområder_bolig	*housing, social (housing), residential areas, house*
Immigration	statsborgerskab_flygtninge_asylansøgere_asyl	*citizenship, refugees, asylum-seekers, asylum*
Infrastructure	dsb_trafik_digital_lufthavn	*DSB, traffic, digital, airport*
Justice	politiet_kriminalitet_hunde_fængsel	*police, crime, dogs, prison*
Labour	arbejdsmarkedet_job_pension_pensionsalderen	*labour market, job, pension, retirement age*
Politics	statsministeren_ministeren_minister_folkeparti	*the prime minister, the minister, minister, people's party*
Regional Affairs	kommunerne_kommuner_kommunernes_kommune	*municipalities*
Social Affairs	kvinder_ældre_mænd_ældreplejen	*women, elders, men, elderly care*
Territories	grønland_grønlandske_færøerne_arktis	*Greenland, Greenlandic, The Faroe Islands, the artic*

Appendix 2: Statement Distribution over Time

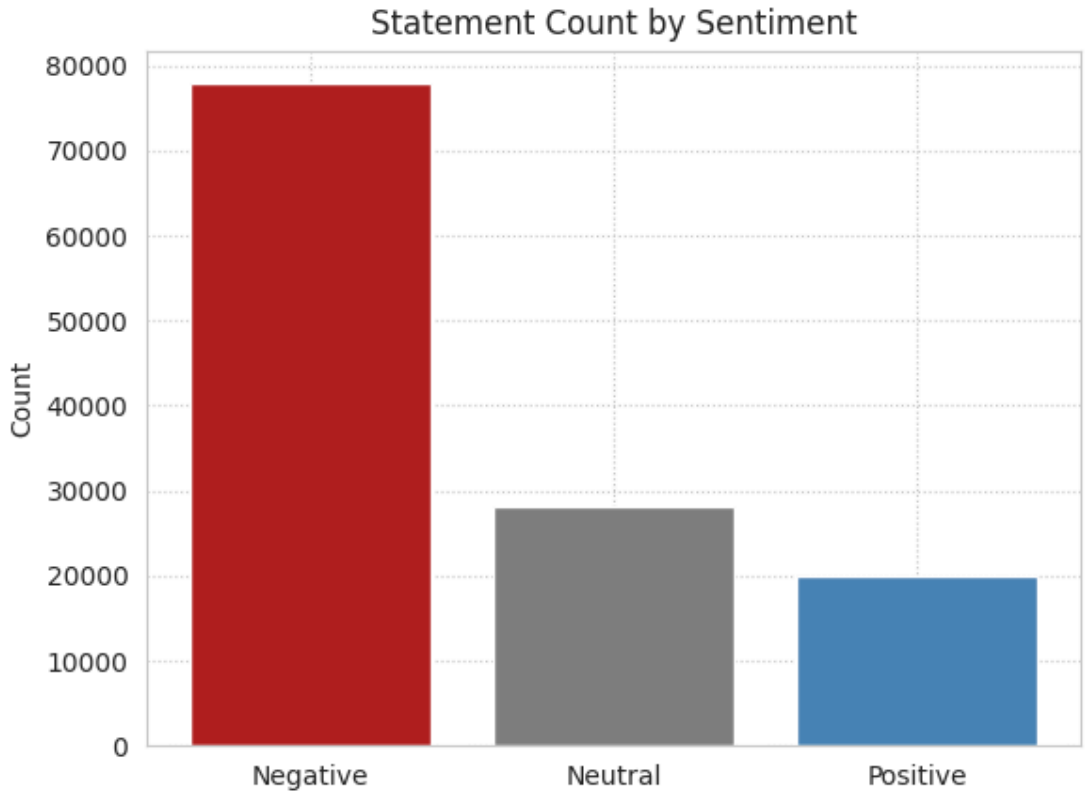


Appendix 3: Statement Distribution by Topic over Time

Appendix 4: Party Statement Distribution

	Count	Percentage
Socialdemokratiet (S)	28089	22,3%
Venstre (V)	20015	15,9%
Enhedslisten (EL)	14272	11,3%
Dansk Folkeparti (DF)	11948	9,5%
Socialistisk Folkeparti (SF)	10291	8,2%
Konsevativt Folkeparti (KF)	9931	7,9%
Danmarksdemokraterne (DD)	8649	6,9%
Radikale Venstre (RV)	8191	6,5%
Liberal Alliance (LA)	6775	5,4%
Alternativet (ALT)	2650	2,1%
Moderaterne	1977	1,6%
Greenland and the Faroe Islands	1660	1,3%
Nye Borgerlige (NB)	1573	1,2%

Appendix 5: Sentiment Distribution



Appendix 6: Translation of Party Names with Corresponding Bloc

Abbreviation	Party Name (Danish)	Party Name (English)	Bloc
S	Socialdemokratiet	The Social Democratic Party	Red *(left)*
V	Venstre	The Liberal Party	Blue *(right)*
EL	Enhedslisten	The Red-Green Unity List	Red *(left)*
DF	Dansk Folkeparti	Danish People's Party	Blue *(right)*
SF	Socialistisk Folkeparti	Socialist People's Party	Red *(left)*
KF	Konservative Folkeparti	Conservative People's Party	Blue *(right)*
DD	Danmarksdemokraterne	Denmark Democrats	Blue *(right)*
RV	Radikale Venstre	Danish Social Liberal Party	Red *(left)*
LA	Liberal Alliance	Libaral Alliance	Blue *(right)*
ALT	Alternativet	The Alternative	Red *(left)*
M	Moderaterne	The Moderates	Blue *(right)*
NB	Nye Borgerlige	New Right	Blue *(right)*
g_f	Grønland og Færøerne	Greenland and the Faroe Islands	*Shifting affiliations*