Government and Opposition in Danish Parliamentary Debates

Costanza Navarretta, Dorte Haltrup Hansen
University of Copenhagen
Emil Holms Kanal 2, 2300 Copenhagen, Denmark
costanza@hum.ku.dk, dorteh@hum.ku.dk
https://nors.ku.dk/english/staff/?pure=en/persons/53191,
https://nors.ku.dk/english/staff/?pure=en/persons/127790

Abstract

In this paper, we address government and opposition speeches made by the Danish Parliament's members from 2014 to 2022. We use the linguistic annotations and metadata in ParlaMint-DK, one of the ParlaMint corpora, to investigate some characteristics of the transcribed speeches made by government and opposition and test how well classifiers can identify the speeches delivered by these groups. Our analyses confirm that there are differences in the speeches made by government and opposition e.g., in the frequency of some modality expressions. In our study, we also include parties, which do not directly support or are against the government, the *other* group. The best performing classifier for identifying speeches made by parties in government, in opposition or in *other* is a transformer with a pre-trained Danish BERT model which gave an F1-score of 0.64. The same classifier obtained an F1-score of 0.77 on the binary identification of speeches made by government or opposition parties.

Keywords: Parliamentary Speeches, Classification, Government/Opposition

1. Introduction

This paper addresses the parliamentary speeches delivered by Danish politicians in government, in opposition or in a group called *other*, which comprises parties neither supporting directly the government nor being against it. More precisely, we want to investigate whether the speeches by the three groups are different in some linguistic aspects, and then we apply classifiers to their transcriptions in order to automatically identify which of the three groups produced the speeches.

The data we use are extracted from ParlaMint-DK, one of the 29 corpora in the ParlaMint v. 4.0[1]. The corpora were collected and annotated under the ParlaMint project[2], which was initiated and partially funded by the European CLARIN infrastructure [3] (Erjavec et al., 2022).

ParlaMint-DK covers the debates of the Danish parliament, *Folketinget*, in the period 2014-2022. As all the other ParlaMint corpora, ParlaMint-DK contains various information types, hereunder the party, gender and age of the speaker, as well as information on whether the party of the speaker at that time is in government or opposition. Moreover, parties that are in neither group (*other*) can be identified.

The ParlaMint corpora also contain automatically produced linguistic annotations in the same theoretic framework. Furthermore, all corpora are encoded in the same TEI format and contain the same type of metadata (Erjavec et al., 2022). The corpora are both available as texts[4] and in a linguistically annotated version[5].

Recently, many of the ParlaMint corpora have been automatically translated into English[6].

The paper is organized as follows. In section 2, we shortly present some background studies, and in section 3 we describe the data and account for some linguistic differences in the speeches made by government, opposition and *other*. In section 4 we outline related work on automatic text classification, and in section 5, we present our classification experiments. Finally, in section 6, we discuss our results and in section 7 we conclude and outline future work.

2. Background Studies

Various researchers have addressed the speeches made by government and opposition parties. Many of these studies focus on different aspects related to the sentiment expressed in the speeches by the two groups, see e.g. the overview in (Abercrombie and Batista-Navarro, 2020).

For example, Sawhney et al. (2020) address the automatic identification of the political stance in speeches by government and opposition par-

[1] All the corpora are freely available at https://www.clarin.si/repository/xmlui/handle/11356/1864.

[2] https://www.clarin.eu/parlamint

[3] https:\clarin.eu

[4] https://www.clarin.si/repository/xmlui/handle/11356/1859

[5] https://www.clarin.si/repository/xmlui/handle/11356/1860

[6] https://www.clarin.si/repository/xmlui/handle/11356/1864

ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora, pages 154–162

ties, while Curini et al. (2020) analyse government and opposition in the Japanese parliament over sixty years using Wordfish, a method which uses a scaling technique for predicting positions based on word frequencies in political texts (Slapin and Proksch, 2008).

Izumi and Medeiros (2022) apply a Naive Bayes Classifier to the speeches in the Brazilian Senate in order to classify the positive or negative sentiment presented by the speakers when talking on different issues. The authors annotated a number of speeches manually in order to train and test the classifier. They find that the differences in sentiment between the speeches do not correspond to the left-wing and right-wing dichotomy as they expected, but they reflect much more the government and opposition division. In their opinion, this result indicates that the politicians in the government use a more sentiment rich language in order to influence the politicians in the Senate to vote in favor of their bills.

In our study, we were partly inspired by the findings in (Izumi and Medeiros, 2022). Differing from their work, however, we do not look at the sentiment expressed by politicians , but we use linguistic features of the transcriptions of the Danish speeches in order to determine whether the speeches made by the three groups *government*, *opposition* and *other* differ and can, therefore, be automatically identified, even if the policy stances of many Danish parties are common in many cases, at least with respect to how they vote in the parliament. More precisely, most Danish parties collaborate during the various legislative periods, and many laws are therefore supported by both parties in government and parties outside it. In fact, counting the votes in the parliament in the investigated period, we found that in approx. 22% of the cases the votes were unanimous. Moreover, the two parties, *The Social Democratic Party* and *The Liberal Party*, which belonged to opposite wings and chaired each two of the governments in this period, also expressed the same votes in additionally 8.2% of the cases. This means that in more than 30% of the cases, the politicians of the two main parties voted in the same way independently on whether they were in government or opposition.

3. The Data

The data in our studies was extracted from the annotated ParlaMint-DK, one of the annotated ParlaMint v. 4.0 corpora[7]. The ParlaMint-DK corpus covers the transcriptions of the speeches from the 7 October 2014 to the 7 June 2022. The transcriptions and some of the metadata included in ParlaMint-DK were downloaded from the Danish Parliament website[8], while other metadata and the linguistic annotations of the corpus were made by researchers from CLARIN-DK (Jongejan et al., 2021).

Speech Group	Speeches	Tokens
Government	56,369	14,039,122
Opposition	74,922	16,919,425
Other	59,403	13,542,700
Speaker	207,915	3,139,068
Total	398,609	47,640,315

Table 1: Number of speeches and tokens in the corpus.

ParlaMint-DK contains 47,640,315 tokens, 3,139,068 uttered by the Speaker (the chair) and 44,501,247 tokens uttered by the members of the parliament and the ministers. Only the latter speeches are relevant for this work. All these speeches are marked as either belonging to the government, the opposition or none of the two (the *other* group).

The government can comprise one or more parties; the opposition always consists of more parties from the opposite political wing in the studied period. The group *other* is more heterogeneous. It consists of both parties which give parliamentary support to the government, without being part of it, and parties which are not in direct opposition to the government. *other* also comprises small independent parties, e.g., the parliament members from Greenland and the Faroe Islands, which have acted as parliamentary support to various governments.

The distribution of the speeches in the three groups, and the number of tokens in them are in Table 1.

The Speaker often takes the floor, but does not speak for a long time since the Speaker's role is to chair the meetings and ensure that the formal rules are followed (average number of tokens per speech is 11). The largest number of speeches comes from the opposition parties, followed by the *other* parties. The government parties take the floor less often than the parties in the two other groups, but their speeches are longer (in average 249 words per speech) than the speeches made by the *other* parties (228 words per speeches) and the opposition parties (226 words per speeches). The fact that members of the government parties speak for a longer time than those in the other groups is not surprising since ministers often present the bills.

There were 20 parties in the Danish parliament in the investigated time span. Table 2 shows the positions of at the time largest 11 left and right-wing

[7]The corpus is available a from https://www.clarin.si/repository/xmlui/handle/11356/1864 as the other ParlaMint annotated corpora.

[8]ftp://oda.ft.dk

	Government	Opposition	Other
Left w.		EL	EL
		SF	SF
		ALT	ALT
	S	S	
	RV	RV	RV
Right w.			M
	V	V	
	KF	KF	KF
	LA	LA	LA
			DF
			NB

Table 2: Largest parties' positions in the investigated period.

parties in the various legislative periods., that is some parties were always in the *other* groups, while some parties in some legislation periods were in government, while in other ones were in opposition. The remaining 9 parties, all part of the group *other* are smaller, they have never been in a government, and their members seldom take the floor. They are not shown in Table 2. The 11 parties shown in Table 2 from the left to the right are the following:

EL The Red-Green Unity List (*Enhedslisten*)

SF Socialist People's Party (*Socialistik Folkeparti*)

ALT The Alternative (*Alternativet*)

S The Social Democratic Party (*Socialdemokratiet*) has been leading two governments in the investigated period (2014-2016, and 2019-)

RV Danish Social Liberal Party (*Radikale Venstre*)

V The Liberal Party (*Venstre*) has been leading two right-wing governments in the investigated time (2009-2014, 2016-2019)

K Conservative People's Party (*Konservative Folkeparti*)

LA The Liberal Alliance (*Liberal Alliance*)

DF Danish People's Party (*Dansk Folkeparti*)

NB New Right (*Nye Borgerlige*)

In the period covered by the ParlaMint-DK data, the Social Democrats (S) and the Liberals (V) are always either in government or in opposition, while other parties like The Red/Green Alliance (EL) or Danish People's Party (DF), are never part of the government. In the Lars Løkke Rasmussen II Cabinet (28.06.2015 - 28.11.2016), the government consisted of only one party, The Liberal Party (V), while Danish People's Party (DF), Liberal Alliance (LA) and Conservative People's Party (KF) were the parliamentary support. From 28.11.2016 to 27.06.2019, the liberals (V) were at the government with the Liberal Alliance (LA) and Conservative People's Party (KF). The opposition consisted of the left-wing parties, which also comprised a centre party, the Danish Social Liberal Party. From 2014 to 28.06.2015 the social democrats (S) headed a left-wing government which also comprised ministers from the Danish Social Liberal Party (RV). After the election in 2019, in the Mette Frederiksen I Cabinet (27. 06 2019 til 15. 12 2022), the social democrats alone formed the government with the other "left-wing" parties as parliamentary support. During these governments, the right-wing parties were the opposition.

3.1. Analysis of the Speeches

The data from the ParlaMint-DK annotated corpus, which we use in the present research are the following: the tokenised transcriptions, the lemmatised transcriptions and, for each speech, information about whether it was delivered by a speaker whose party was in government (GOV), in opposition (OPPN) or in the *other* group.

In our first study, we looked into whether there is an overlap of the lemmas in the three groups of speeches, and we found that 60,989 lemmas only occurred in the government speeches, 13,225 only occurred in the opposition speeches and 34,333 lemmas only occurred in the speeches by the *other* parties. Thus, we found that the government speeches contained the largest number of lemmas which did not appear in the speeches of the other groups, followed by the speeches of the *other* group. A first analysis of the lemmas that only occur in each of the three groups indicates that they mostly consist of compounds, such as *affaldshåndteringsgebyr* (waste management fee), which only occurs in the speeches made by parties in government and *affaldsforbrændingskapacitet* (waste incineration capacity) which only occur in the speeches by opposition parties. This indicates that even if the topics discussed in the parliament by the parties in the three groups are the same, the politicians can address different details about the same topics. Moreover, the data shows the great amounts of compounds which characterise Danish as other Germanic languages.

In the second study, we wanted to investigate the speaker's attitudes to what is said by looking into some of the ways of expressing modality in the speeches. The use of modality in political speeches has been addressed in several studies since through modality speakers can express their attitudinal state towards what they say or others have expressed, see e.g., (Simon-Vandenbergen, 1996; Lillian, 2008; Sharififar and Rahimi, 2015).

The most frequent way of expressing modality

in Danish is with modal auxiliaries and modal adverbs. More specifically, *mood* in verbs usually expressed the speaker's or another person's attitude towards an utterance, e.g., (Allan et al., 2015),. The modal auxiliaries in Danish are *kunne* (could), *skulle* (should), *ville* (would), *måtte* (had to), *turde* (dare), *burde* (ought to), *gide* (bother). When they are used in past tense, they often indicate a non-factual (hypothetical) attitude to what is said, while when they are used in present tense, they often indicate a firmer and more factual attitude.

Examples of the modal auxiliary *skulle* in 1) present tense and 2) past tense, are the following:

1. S: *jeg **skal** som med de foregående dobbeltbeskatningsoverenskomster også meddele at Socialdemokratiet støtter dette lovforslag*
   (I **must** also announce like with the previous double taxation agreements that the Social Democracy supports this bill)

2. V: *det var bare lige for at notere at vi også gerne stadig væk **skulle** have en positiv stemning i frikommunerne*
   (it was just to note that we still *would like* to maintain a positive atmosphere in the free municipalities)

In the first example, a social democrat in government presents the position of its party with respect to the existing double taxation agreements (a fact), while in the second example, a liberal in the opposition express a desire.

Danish modal adverbs are divided by Jensen (1997) into epistemic and factual adverbs. As for the modal auxiliaries, the distinction between the two groups is that the epistemic adverbs can indicate a more hesitant attitude, while the factual adverbs show more firmness. The epistemic adverbs listed in (Jensen, 1997) are the following: *måske* (maybe), *nok* (probably), *muligvis* (possibly), *dog* (though), *vist* (possibly), *formodentlig* (probably), *åbenbart* (apparently), *tilsyneladende* (seemingly), *egentlig* (actually), *vel* (I guess), while the factual adverbs are *desværre* (unfortunately), *uheldigvis* (unfortunately), and *heldigvis* (fortunately).

Examples of 1) a factive adverb and b) an epistemic adverb are in what follows:

1. V: ***heldigvis** er der flere unge med minoritetsbaggrund, der blander sig i debatten og siger fra*
   (**fortunately**, there are several young people with minority backgrounds, who are getting involved in the debate and put their foot down)

2. EL: *hvis ikke det her lovforslag, som **muligvis** krænker menneskerettighederne, og som i hvert fald træder på retssikkerheden, blev vedtaget*
   (if this bill, which **possibly** violates human rights and certainly undermines legal certainty, was not adopted)

Group	**Modal pres**	**Modal past**
Government	480,687	60,492
Opposition	552,544	90,020
Other	453,628	77,532
Group	**Factive adv**	**Epistemic adv**
Government	5,412	57,171
Opposition	5,903	80,466
Other	4,658	69,340

Table 3: Occurrences of modal auxiliaries and modal adverbs

Group	**Modal pres**	**Modal past**
Government	3.58	0.44
Opposition	3.49	0.57
Other	3.49	0.6
Group	**Factive adv**	**Epistemic adv**
Government	30.39	0.42
Opposition	0.37	0.51
Other	0.36	0.53

Table 4: Relative frequency of modal auxiliaries and modal adverbs

In the first example a liberal expresses a fact, while in the second a example member of the Red-green Union list expresses a possibility regarding a bill, which might violate human rights. We extracted the two types of modal auxiliary verb (present vs. past form) and the factual vs. epistemic adverbs in the parliamentary speeches by government, opposition and *other* group in order to determine whether the parties in government use more confident and factual expressions, and the parties in the other two groups express less confidence when they speak as e.g., was noted in the sentiment analysis of the speeches made by the politicians in the Brazilian senate (Izumi and Medeiros, 2022).

In table 3, the number of each type of modal auxiliary and clausal adverb in each group of speeches is shown, while table 4 shows their relative frequency.

There are no statistically significant differences in the occurrences of modal auxiliaries in present tense and of factual adverbs in the speeches by the three groups. On the contrary, we found significant differences (chi-square's $p < 0.0001$, $df = 1$) in the use of both past tense modal auxiliaries and epistemic adverbs in the speeches by politicians in government and politicians in the other two groups. The politicians in government use significantly less epistemic adverbs and non-factual modal auxiliaries than the politicians in opposition or in the *other* group, thus the politicians in government express themselves in a more confident way.

We also investigated whether we could find the same differences in the speeches of politicians belonging to the two parties that chaired left-wing and right-wing governments (the Social Democrats and the Liberals) comparing cases when they were chairing the government and when they were in opposition, and the above differences in the use of modal auxiliaries in past tense and of epistemic adverbs were confirmed with the same significance values.

These results show that politicians in government use less hypothetical constructions than the politicians that are not in government.

Concluding, our first quantitative study indicates that there are differences in the speeches by the three groups' politicians, and the second study shows differences between speeches made by government parties and parties not in the government. These results are promising for applying text classification to the transcriptions.

4. Text Classification: Related Work

Automatic text classification is one of the main applications of natural language processing. It aims to assign pre-defined labels to whole texts or parts of them. Machine learning based approaches use annotated data to identify the labels in non-annotated data.

The features and algorithm that have been tested the past decades are many, see e.g., (Kowsari et al., 2019; Minaee et al., 2021). The most frequently used features are n-grams, word vectors, TF*IDF (Term Frequency * Inverse Document Frequency)[9] vectors, word embeddings (Kowsari et al., 2019).

Traditional machine learning classifiers comprise e.g., Naïve Bayes and Logistic regression, while examples of deep learning methods used for classification are Multilayer Perceptrons (MLP), Recurrent Neural Networks (RNN), and Long-Short Term Memory systems (LSTM). More recently transformers and pre-trained large language models have improved the state-of-the-art results on some of the most common classification tasks such as sentiment analysis and classification of news articles (Minaee et al., 2021).

Text classification has also been applied to political data, and specifically to parliamentary debates. Many of these studies have addressed the classification of opinions in the debates, inter alia (Abercrombie and Batista-Navarro, 2018; Sawhney et al., 2020), but also the automatic identification of ideology or position in the speeches (Proksch and Slapin, 2012; Riabinin, 2009), the automatic identification of policy domains (Ristilä and Elo, 2023; Navarretta and Hansen, 2022) and of parties (Kapočiūtė-Dzikienė and Krupavičius, 2014; Navarretta and Hansen, 2020).

In our classification experiments, we follow this line of research with the aim of identifying speeches by government, opposition and *other* parties. We test traditional machine learning classifiers and a neural network classifier training them on the most frequently used representations of the ParlaMint-DK transcriptions. In the final experiments, we applied a transformer and a pre-trained Danish BERT model to our data.

5. Classification Experiments

The aims of our classification experiments were to test to which extent various feature types and machine learning classifiers can predict if speeches are delivered by politicians in *Government*, *Opposition* or *other*.

The data we used were the tokenised and lemmatised transcriptions of the speeches, as well as information about whether the speaker's party was in government, opposition or in the *other* group.

The experiments were run in python 3 and the main libraries used are Pandas, Numpy, and Scikit-learn[10]. For the final experiments with a transformer and pre-trained BERT model, pytorch[11] was used.

Firstly, we ran a number of classifiers on the word vectors and TF*IDF vectors of tokens and lemmas in order to determine whether the former or the latter dataset performed best for this task. All classifiers gave the best results with lemma based features. The classifiers we tested were a stratified classifier[12], which is our first baseline, a Multinomial Naïve Bayes, our second baseline, Logistic Regression[13], and a Multilayer Perceptron Classifier[14]. All classifiers' implementations were those provided in Scikit-learn.

We also ran these classifiers with vector and TF*IDF vector representations of the data's bigrams and trigrams.

Secondly, we ran the classifiers and the unigrams features from the first experiments on speeches from only government and opposition (binary classification), since the government vs opposition distinction is used in many political studies. Moreover, many of the parties in the *other* group only

[9] TF*IDF is a technique proposed in (Luhn, 1958) and then adopted by both information retrieval and NLP. It allows to identify documents on the basis of the frequency of their words relative to the words' frequency in the whole dataset.

[10] https://scikit-learn.org/stable/

[11] https://pytorch.org/

[12] The classifier generates predictions by following the training set's class distribution.

[13] Logistic Regression was run with the *lbfgs* solver.

[14] Mulilayer Perceptron was run with the *sgd* solver, *tanh* activation, $alpha = 0.001$, 3 hidden layers, and *constant* learning rate.

belonged to it in the investigated period, and we wanted to address especially parties that have been part of different groups in different periods in order to be sure that linguistic differences in the speeches are not exclusively party dependent.

Finally, we run on the lemmas of the speeches a hugging face transformer[15] with a pre-trained Danish BERT model[16], which has been trained and distributed by the Danish company Certainly[17].

In the first group of experiments, we tested word vectors and the TF*IDF vectors with 15,000 to 19,000 features. The results of classification improved when going from 15,000 features to 17,000 and then decreased. Therefore, we only report the results obtained with the two vectorized datasets and $max\_features = 17000$. The same number of vector features were then also used in the second group of experiments. 10-fold cross validation was performed and Precision (P), Recall (R) and weighted F1-score (F1) are given as evaluation measures.

The results when the classifiers were trained on unigrams, bigrams and trigrams vector representations are in Table 5.

Naïve Bayes classifier outperforms the stratified baseline (F1-score 0.34 vs. 0.47) and is the only algorithm that performs slightly better when trained on vectors of bigrams and unigrams. Both Logistic Regression and Multilayer Perceptron outperform the second baseline, that is the results of the Naïve Bayes classifier. The best results are also produced by Logistic Regression trained on TF*IDF vectors of lemmas with $F1 = 0.61$. Multilayer Perceptron also performs best when trained on TF*IDF unigrams' vectors. The results of the two classifiers decrease slightly when they were run on the vectorized bigrams, and their performance decreases even more when they were trained on the two types of vectorized trigrams.

The confusion matrix from Logistic Regression trained on the TF*IDF lemma vectors is in figure 1[18].

The classes that are most often confused with each other are *Opposition* and *other*, and this could be expected since they both consist of speeches made by parties that are not in government. We also analyzed some of the erroneously classified speeches and found that some were short, and/or they did not address a specific political issue, as it is shown in the following speech examples:

- *ja* (yes)
- *tak* (thank you)
- *nej det har jeg ikke* (no, I have not)

Classifier	P	R	F1
Stratified	0.34	, 0.34	0.34
Lemma vectorized			
NaïveBayes	0.52	0.49	0.47
LogisticR	0.58	0.58	0.58
MultilayerP.	0.6	0.593	0.594
TF*IDF			
NaïveBayes	0.541	0.49	0.0.47
LogisticR	0.61	0.61	**0.61**
MultilayerP.	0.6	0.6	0.6
Bigrams vectorized			
NaïveBayes	0.52	0.5	0.48
Logistic	0.57	0.57	0.57
MultilayerP.	0.51	0.51	0.51
TF*IDF bigrams			
NaïveBayes	0.53	0.502	0.483
LogisticR	0.6	0.6	0.6
MultilayerP.	0.584	0.584	0.584
Trigrams vectorized			
NaïveBayes	0.513	0.51	0.50
LogisticR	0.533	0.534	0.533
MultilayerP.	0.472	0.472	0.472
TF*IDF trigrams			
NaïveBayes	0.52	0.5	0.48
Logistic	0.553	0.554	0.552
MultilayerP.	0.541	0.542	0.541

Table 5: Results of the first classification experiments

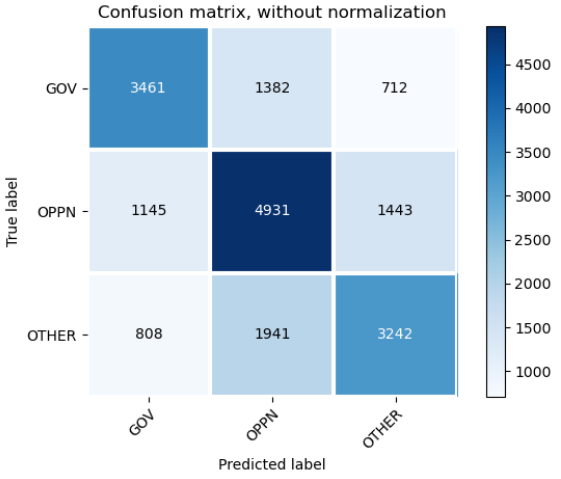

Figure 1: Confusion matrix for ternary classification with Logistic Regression

[15] https://huggingface.co/docs/transformers/index

[16] Version 2, https://github.com/certainlyio/nordic_bert

[17] https://certainly.io/

[18] In the two confusion matrices in the paper, GOV, stands for government, OPPN for opposition, and OTHER for *other* since these were the labels used in the dataset.

Classifier	P	R	F1
Stratified	0.5	0.5	0.5
Lemma vectors			
NaiveBayes	0.654	0.66	0.65
Logistic	0.733	0.734	0.732
MultilayerP.	0.735	0.736	0.737
TFIDF vectors			
NaiveBayes	0.69	0.68	0.66
Logistic	0.752	0.753	**0.751**
MultilayerP.	0.741	0.742	0.741

Table 6: Results of the binary classification experiments

- *jamen så kan man rejse et civilt søgsmål* (well then you can bring a civil action)
- *jeg tror ikke at jeg har yderligere kommentarer* (I do not think that I have further comments)

In the second group of experiments, we applied the same classifiers and used the same unigrams features as in the first group of experiments, but in this case we only addressed the speeches made by parties in government and opposition (binary classification). The results of these experiments are in Table 6.

Also in these experiments, the Multinomial Naïve Bayes classifier outperforms the stratified classifier, and both Logistic Regression and Multilayer Perceptron give better results than the Naïve Bayes classifier, which also performs quite well on this task. Also in these experiments, the best results were achieved by Logistic Regression trained on TF*IDF lemma vectors (F1-score$= 0.751$). The F1-score of Logistic Regression outperforms the F1-score of the Stratified classifier with more than 0.25. Multilayer Perceptron gave slightly better results than Logistic Regression when the two classifiers were trained on word vector representations, while it gave slightly worse results when trained on TF*IDF vectors.

The confusion matrix from the binary classification performed by Logistic Regression trained on the TF*IDF lemma vectors is in figure 2. The confusion matrix shows that speeches made by the government are more often classified as speeches made by the opposition than the contrary. Also in this case, part of the wrongly classified speeches were short and/or did not address a specific political issue.

In the third group of experiments, a bidirectional Encoder Representation from Transformers was run using the pre-trained Danish BERT model[19]. The results for the ternary classification were the

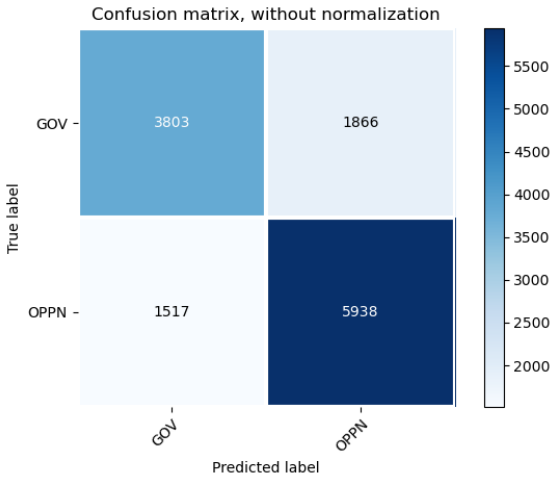

Figure 2: Confusion matrix for binary classification

following: $Precision = 0.68$, $Recall = 0.64$ and F1-score$= 0.64$. The results of the transformer improve especially precision compared to the best results obtained with the more traditional classifiers, but also recall gets better.

Using even larger language models would probably give even better results. However, environmental sustainability issues should be considered, since it took much more time to fine tune and train the transformer on this data than training Logistic Regression (48 hours vs. half an hour) even if we used a much stronger processor when running the transformer than when training Logistic Regression.

Finally, we run the transformer and the pre-trained Danish BERT model on the data consisting only of speeches made by government and opposition parties. The data was so that 80% was used for fine tuning the pre-trained model and 10% were used for testing and 10% for validation.The results for the binary classification were the following: $Precision = 0.79$, $Recall = 0.77$ and F1-score$= 0.77$. Also in this case, the transformer gives the best results.

6. Discussion

Our first quantitative analyses of the parliamentary speeches in the ParliaMint-DK corpus show that there are differences in the speeches delivered by government, opposition or the *other* group.

The politicians in the government use less hypothetical constructions than politicians in the other two groups. Moreover, the fact that a number of lemmas in the speeches of each group do not occur in the speeches produced by politicians belonging

[19]The experiment was run on an Intel Xeon gold processor with 64 cores and 364 GB memory provided by https:cloud.sdu.dk. Optimization was performed with the pytorch implementation of the AdamW optimizer. The learning rate was $5e-5$ and $eps = 1e-8$ (the default). 16 batches and 4 epochs were used.

to the other two groups might indicate that there are issues, which are addressed more by one group or that politicians in government, opposition and *other* parties use some particular words depending on their party's current position.

This aspect should be examined further. In future, we could also investigate whether the differences between the three groups are more evident when they address specific policy areas.

The results of our ternary classification experiments (F1-score$= 0.64$) confirm that identifying the speeches of politicians in government, opposition and outside the two groups are quite good given the type of data. The best results were obtained with a transformer trained on a BERT, but also a traditional ML classifier, Logistic Regression, trained on TF*IDF vectors of lemmas gave a good F-score (0.61).

The results of ternary classification when traditional ML classifiers were trained on bigrams and trigrams vector representations gave different results depending on the classifier and the type of vector, but in general the results decreased slightly when going from unigrams to bigrams, and even more when trigrams were used.

In our binary classification experiments, we again obtained the best results using the transformer and the pre-trained Danish BERT model, with an F1-score of 0.77. This result is also good when compared to the results obtained by other researchers on different text classification tasks (Minaee et al., 2021). The second best result was again obtained by Logistic Regression on TF*IDF vectors of lemmas (best results with 17,000 features: $F1 - score = 0.754$). The analysis of randomly selected speeches, which were wrongly classified, showed that some of them were short and did not address a specific policy domain. Many of these examples, in fact, had a communication management function (Bunt et al., 2010).

7. Conclusions and Future Work

In this paper, we have presented quantitative analyses of the transcriptions of Danish parliamentary speeches as well classification experiments aimed to determine whether the speeches were produced by politicians in government, opposition or *other* parties. Both the results of our preliminary analyses of the speeches and our ternary and binary classification experiments show that there are differences between the speeches of government parties and parties outside it. These differences were also found within parties taking either the role of chairing the government or being in opposition in different years of the investigated period.

The results of this study also confirm some of the observations by Izumi and Medeiros (2022) who classified sentiment in Brazilian Senate speeches delivered by left-wing and right-wing parties.

Future extensions of our work are many, such as a) making further analyses of the linguistic characteristics of the speeches of government parties and parties outside the government, b) investigating whether there are policy domains which are more often addressed by each of the three groups, c) reducing the classification experiments to the speeches of one of the two large parties which have been in government and in opposition in different periods, and d) comparing the results from this study with similar studies of the speeches from other ParlaMint corpora. Since all the ParlaMint corpora have the same metadata and linguistic annotation types (Erjavec et al., 2022), it should be possible to extend this kind of study to other parliamentary data also comparing language specific characteristics of e.g., speeches made by government and opposition parties. Moreover, the English translation of ParlaMint-DK could be used in a replication study in order to evaluate the quality of the automatic translation.

Finally, more Large Language Models could be tested for classification, but environmental sustainability should be considered given the larger amount of resource they require compared with traditional machine learning classifiers.

8. Acknowledgements

We would thank the CLARIN infrastructure for supporting the ParlaMint project and making the ParlaMint data available.

Part of the computation done for this article was performed on the UCloud interactive HPC system, which is managed by the eScience Center at the University of Southern Denmark.

9. Bibliographical References

Gavin Abercrombie and Riza Batista-Navarro. 2020. Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, 3(1):245–270.

Gavin Abercrombie and Riza Theresa Batista-Navarro. 2018. Identifying opinion-topics and polarity of parliamentary debate motions. In *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 280–285.

Robin Allan, Tom Lundskaer-Nielsen, and Philip Holmes. 2015. *Danish: A Comprehensive Grammar*, first edition. Routledge, London.

H. Bunt, J. Alexandersson, J. Carletta, J.-W. Choe, A. Chengyu Fang, K. Hasida, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, and D. Traum. 2010. Towards and iso standard for dialogue act annotation. In *Proceedings 7th international conference on language resources and evaluation (LREC 2010)*, pages 2548–2555.

Luigi Curini, Aito Hino, and Atsushi Osaka. 2020. The Intensity of Government–Opposition Divide as Measured through Legislative Speeches and What We Can Learn from It: Analyses of Japanese Parliamentary Debates, 1953–2013. *Government and Opposition*, 55(2):184–201.

Tomaž Erjavec, Maciej Ogrodniczuk, Petya Osenova, Nikola Ljubešic, Kiril Simov, Andrej Pancur, Michał Rudolf, Matyáš Kopp, Starkadhur Barkarson, Steinthór Steingrímsson, Çagrı Çöltekin, Jesse de Does, Katrien Depuydt, Tommaso Agnoloni, Giulia Venturi, María Calzada Pérez, Luciana D. de Macedo, Costanza Navarretta, Giancarlo Luxardo, Matthew Coole, Paul Rayson, Vaidas Morkevicius, Tomas Krilavicius, Roberts Dargis, Orsolya Ring, Ruben van Heusden, Maarten Marx, and Darja Fiser. 2022. The ParlaMint corpora of parliamentary proceedings. *Language Resources and Evaluation*.

Mauricio Y Izumi and Danilo B Medeiros. 2022. Government and opposition in legislative speechmaking: Using text-as-data to estimate brazilian political parties' policy positions–corrigendum. *Latin American Politics and Society*, 64(1):174–175.

Eva Skafte Jensen. 1997. Modalitet og dansk. *NyS, Nydanske Sprogstudier*, 23(23):9–24.

Bart Jongejan, Dorte Haltrup Hansen, and Costanza Navarretta. 2021. Enhancing CLARIN-DK Resources While Building the Danish ParlaMint Corpus. In *CLARIN Annual Conference 2021 Proceedings*, pages 70–73. CLARIN ERIC.

Jurgita Kapočiūtė-Dzikienė and Algis Krupavičius. 2014. Predicting party group from the Lithuanian parliamentary speeches. *Information Technology and Control*, 43(3):321–332.

Kamran Kowsari, Kiana Jafari Meimandi, Mojtaba Heidarysafa, Sanjana Mendu, Laura Barnes, and Donald Brown. 2019. Text classification algorithms: A survey. *Information*, 10(4):150.

Donna L Lillian. 2008. Modality, persuasion and manipulation in canadian conservative discourse. *Critical Approaches to Discourse Analysis across Disciplines*, 2(1):1–16.

Hans Peter Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165.

Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. Deep learning–based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 54(3):1–40.

Costanza Navarretta and Dorte Haltrup Hansen. 2020. Identifying parties in manifestos and parliament speeches. In *Proceedings of the Second ParlaCLARIN Workshop*, pages 51–57.

Costanza Navarretta and Dorte Haltrup Hansen. 2022. The Subject Annotations of the Danish Parliament Corpus (2009-2017) - Evaluated with Automatic Multi-label Classification. In *Proceedings of LREC 2022*. ELRA.

Sven-Oliver Proksch and Jonathan B. Slapin. 2012. Institutional foundations of legislative speech. *American Journal of Political Science*, 56(3):520–537.

Yaroslav Riabinin. 2009. Computational identification of ideology in text: A study of canadian parliamentary debates. *MSc paper, Department of Computer Science, University of Toronto*.

Anna Ristilä and Kimmo Elo. 2023. Observing political and societal changes in Finnish parliamentary speech data, 1980–2010, with topic modelling. *Parliaments, Estates and Representation*, pages 1–28.

Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. 2020. GPolS: A contextual graph-based language model for analyzing parliamentary debates and political cohesion. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4847–4859, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Massoud Sharififar and Elahe Rahimi. 2015. Critical discourse analysis of political speeches: A case study of obama's and rouhani's speeches at un. *Theory and Practice in Language studies*, 5(2):343.

Anne-Marie Simon-Vandenbergen. 1996. Image-building through modality: the case of political interviews. *Discourse & Society*, 7(3):389–415.

Jonathan B Slapin and Sven-Oliver Proksch. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.