Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 1428–1436
Marseille, 20-25 June 2022

The Subject Annotations of the Danish Parliament Corpus (2009-2017) - Evaluated with Automatic Multi-label Classification

Costanza Navarretta and Dorte Haltrup Hansen
University of Copenhagen
Emil Holms Kanal 2, Copenhagen S
costanza@hum.ku.dk, dorteh@hum.ku.dk

Abstract

This paper addresses the semi-automatic annotation of subjects, also called policy areas, in the Danish Parliament Corpus (2009-2017) v.2. Recently, the corpus has been made available through the CLARIN-DK repository, the Danish node of the European CLARIN infrastructure. The paper also contains an analysis of the subjects in the corpus, and a description of multi-label classification experiments act to verify the consistency of the subject annotation and the utility of the corpus for training classifiers on this type of data. The analysis of the corpus comprises an investigation of how often the parliament members addressed each subject and the relation between subjects and gender of the speaker. The classification experiments show that classifiers can determine the two co-occurring subjects of the speeches from the agenda titles with a performance similar to that of human annotators. Moreover, a multilayer perceptron achieved an F1-score of 0.68 on the same task when trained on bag of words vectors obtained from the speeches' lemmas. This is an improvement of more than 0.6 with respect to the baseline, a majority classifier that accounts for the frequency of the classes. The result is promising given the high number of subject combinations (186) and the skewness of the data.

Keywords: Parliamentary Records, Corpus Annotation, Multi-label Classification

1. Introduction

The interest in analysing and automatically processing large amounts of political data has increased in the past decades, and the research has been made possible by the digital availability of political data. One type of political material consists of political speeches from parliaments. These have been the object of investigation in common initiatives, such as the ParlaClarin workshops (Fišer et al., 2018; Fišer et al., 2020), and large data collections, such as the EuroParl parallel corpus (Koehn, 2005) and the ParlaMint corpora (Erjavec et al., 2022). The classification of political discourse in general political subjects, also known as policy areas or domains, has been addressed by researchers in especially political sciences for many decades since subject classifications facilitate the analysis and comparison of how politicians from different political wings address specific subjects nationally or internationally, and over time.

In the present work, we follow this line of research. More specifically, we address the annotation of political subjects in the speeches of the *Danish Parliament Corpus (2009-2017) v2, with subject annotations*, which was released in CLARIN-DK in spring 2021 (Hansen and Navarretta, 2021)[1]. The paper describes how the political subjects were semi-automatically annotated in the corpus using the meetings' agenda titles. The paper also presents an analysis of the annotations and an evaluation of the consistency of the annotations of co-occurring subjects performed through supervised multi-label classification experiments.

The paper is organized as follows. In section 2, background work on the construction of subject annotated political data is shortly presented, then in section 3, our corpus is described, and in section 4, the subject annotations are analysed. In section 5, the multi-label classification experiments performed on the data are accounted for, while the results of the experiments are discussed in section 6. Finally in section 7, we conclude and present future work.

2. Background Work

Different classification schemes have been proposed for organizing political subjects. The most known schemes have been developed in three projects, the Policy Agendas Project[2], the Comparative Agendas Project[3], and the Comparative Manifesto Project[4]. The annotation schemes produced in these projects are described in the form of so-called codebooks.

The scheme by the Policy Agendas Project (PAP) was built in order to structure policy data from various sources and successively investigate changes in the U.S. policy agendas and public policy outcomes starting from the Second World War (Baumgartner et al., 2011).

The scheme created by the Comparative Agendas Project (CAP) aimed to describe the policy activities in more countries. It reuses to large extent the PAP scheme, but it is modified in order to distinguish policy activities around the world. However, it is still biased towards the USA's political system. The CAP

[1] https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/44.

[2] https://liberalarts.utexas.edu/government/news/feature-archive/the-policy-agendas-project.php

[3] https://www.comparativeagendas.net/

[4] https://manifesto-project.wzb.eu/

scheme comprises 21 main subject categories and 192 sub-categories.

The third scheme was built by the Comparative Manifesto Project (CMP) for classifying party election programs (manifestos) and support the investigation of policy preferences of political parties in different countries. It consists of 560 categories.

The two most general classification schemes, CAP and CMP, are too fine-grained with respect to the topics addressed by most parliaments. For this reason, Zirn (2014) and Zirn et al. (2016) propose to use the names of ministries and areas of responsibilities instead of general predefined classes for covering policy activities in the German Parliament. As a result, the speeches of the German Parliament were annotated with 22 subject classes, which correspond to the parliament committees (Zirn et al., 2016).

In Denmark, researchers in political science from the University of Aarhus have manually annotated political data from 1953 to 2007 in the Danish Policy Agendas Project[5] using an adapted version of the Policy Agendas Codebook, and a revised version of the CAP scheme. The data addressed in the Danish Policy Agendas Project comprises different types of material, such as legislative hearings, speeches by the prime minister, debates in parliamentary committees, in the parliament and in city councils.

We have followed the suggestions given by the Danish Policy Agendas Project by relating in the adopted scheme policy activities in the parliamentary debates to the CAP classes. Moreover, we have used classes that correspond to the responsibility areas of the Danish Parliament's committee, inspired by the approach proposed for German parliamentary debates by Zirn (2014) and Zirn et al. (2016). This allows to connect policy activities and the corresponding spokespersons in the Parliament (Hansen et al., 2019).

3. The Danish Parliament Corpus (2009-2017) v.2

The Danish Parliament Corpus (2009-2017), v.2 (DPCv2 henceforth) contains the transcripts of parliamentary speeches of the Danish Parliament (Folketinget) from 6/10-2009 to 7/9-2017. They were downloaded from the Danish Parliament's website[6]. The transcripts are verbatim transcriptions of the speeches, but they have been slightly edited before being released so that the speeches have a colloquial and syntactically coherent written form. Moreover, factual errors and slips of the tongue are corrected in them and spoken language's phenomena such as self-corrections and pauses of various type have not been included (Hansen et al., 2018).

The transcripts from the Danish parliament's web side are also enriched with information about the speaking members of the parliament (their name, role, title and party), and the timing of the speeches. We have added age and gender of the speakers to the original data using external sources, and we have annotated the subjects of the speeches semi-automatically, starting from the manual annotation of the meetings' agenda titles.

By using the agenda titles of meetings to determine the meetings' policy areas, we follow the strategy proposed and adopted by political science researchers in the Policy Agenda Project and in the Comparative Agenda Project.

The DPCv2 consists of 40,841,226 words, 381,949 speeches, 886 files, corresponding to 886 meetings, and 8 zipped folders, each covering a parliamentary year. The files are tab separated text files, one speech per line. The contributions of the Speaker that precede and follow each speech, e.g. *The next speaker is ..."* and *Thanks to Mr. ...*) do not refer to a policy domain and are therefore assigned a null subject. Excluding these contributions, which are not relevant to the present work, we have 183,114 speeches and 38,808,560 words.

3.1. The Annotation Scheme

The annotation scheme comprises 19 main classes that, as mentioned earlier, are a subset of the CAP classification scheme used by Danish political science researchers. The areas of responsibilities in the Parliament (spokesmanships) guided the scheme. This resulted in the inclusion of the categories *Technology*, *Transport* and *IT* all merged into one class, *Infrastructure*.

The scheme consists of the following 19 classes: *Agriculture*, *Business*, *Culture*, *Defence*, *Economy*, *Education*, *Energy*, *Environment*, *European Integration*, *Foreign Affairs*, *Health Care*, *Housing*, *Infrastructure*, *Immigration*, *Justice*, *Labour*, *Local and Regional Affairs*, *Social Affairs* and *Territories*.

Table 1 shows the Danish specific subjects, the corresponding areas of responsibility related to them in the Danish parliament, the CAP codes and areas.

3.2. The Annotation Method

All speeches in the Danish Parliament Corpus (2009-2017), v1 have been recently annotated assigning max. two subjects to each speech. The corpus with subject annotation was released as DPCv2.

The annotations were semi-automatically added to the speeches using the title of the agenda items for the meetings. The method we have followed was described in (Hansen et al., 2019) who annotated a subset of the Danish Parliament Corpus in a pilot study. The method consists in the following steps: a) extracting the titles of the agendas, b) normalizing them, e.g. "First reading of bill 193: XYZ" becomes "XYZ", c) manually annotating the agenda titles with up to two subjects, and d) adding these subjects automatically to each speech under the agenda title.

[5] http://www.agendasetting.dk/.

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

Subject	Area of Responsibility	CAP no.	CAP Areas
Economy	Finance, Fiscal Affairs	1	Domestic Macroeconomic Issues
Health Care	Psychiatry, Health	3	Health
Agriculture	Animal Welfare, Fisheries, Food, Agriculture Consumer Policy	4 1525	Agriculture Consumer Policy
Labour	Labour market	5	Labour and Employment
Education	Higher Education and Research Education	6	Education
Environment	Environment	7	Environment
Energy	Energy Climate	8 705	Energy Air and Noise Pollution, Climate Change and Climate Policies
Immigration	Immigration and Integration, Alien Affairs, Naturalization	9	Immigration and Refugee Issues
Infrastructure	Transportation IT, Media	10 17	Transportation Space, Science, Technology and Communications
Justice	Legal Affairs Constitutional Matters	12 20	Law, Crime, and Family Issues Government Issues
Social Affairs	Children, Family, Social Services, Senior Citizens Gender Equality	13 2	Social Welfare Civil Rights, Minority Issues, and Civil Liberties
Housing	Housing	14	Community Development & Housing Issues
Local and Regional Affairs	Rural Districts and Islands Municipal Affairs	4 2001	Community Development & Housing Issues Local Government Issues
Business	Trade and Industry	15	Industrial and Commercial Policy
Defence	Defence	16	Defence
Foreign Affairs	Foreign Affairs, Development, Cooperation	19	International Affairs and Foreign Aid
European Integration	EU	1910	International Affairs and Foreign Aid
Territories	Faroe Islands, Greenland	2105	Dependencies and Territorial Issues
Culture	Cultural Affairs Ecclesiastical Affairs Sport	23 210 1526	Cultural Policy Issues The Danish national church Sport and Gambling

Table 1: Subjects and the corresponding responsibility areas, CAP numbers, and CAP areas

The speeches that were annotated with two subjects (nearly 5000) in the pilot study were also annotated independently by two other coders who took into consideration the speeches' content. No errors were found in the annotation of the subjects from the normalized agenda titles, but in a few cases (seventeen), the coders disagreed on the order of the main and secondary subject, and proposed to switch them. This was implemented when the coders agreed upon the suggested changes.

4. Analysis of the Annotations

183,114 speeches are annotated with a subject, and 60,387 of these (33%) are also annotated with a second subject. In Figure 1 the distribution of the main subjects in the corpus is given. The most frequently debated subjects in the Danish parliament are *Economy*, followed by *Justice*, *Labour*, *Social Affairs* and *Immigration*. The less frequent subjects are *Territories*, *EU integration*, *Defence*, *Housing*, and *Culture*.

There are 190 combinations of co-occurring subjects, and some of them are frequent, while other combinations only occur few times. Figure 2 shows the 20 most frequently co-occurring subjects. The subject that most often is debated with other subjects is *Economy*. This is not surprising since the economic aspect is essential for many political issues. The subjects that are more frequently debated together are the following: a) *Economy* and *Local and regional affairs*, b) *Environment* and *Agriculture*, c) *Economy* and *Business*, d) *Labour*

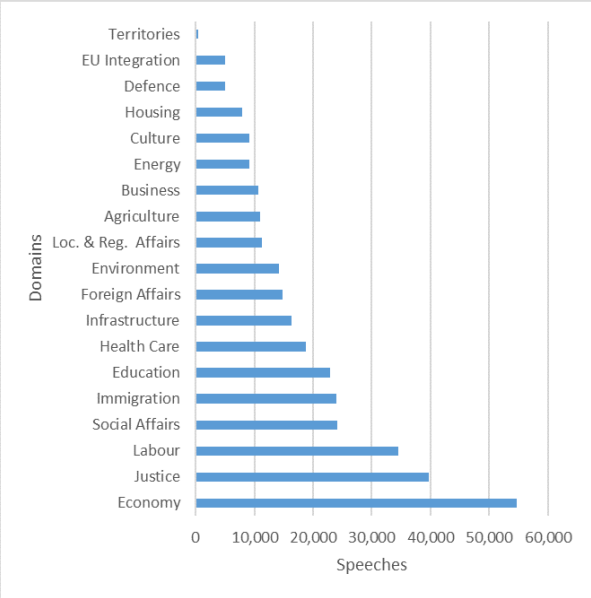

Figure 1: The distribution of the main subjects

and *Social affairs*. Some of the rarely co-occurring subjects in our data are a) *Health Care* and *Justice*, b) *Education* and *Housing*, c) *Defence* and *Economy*, e) *Business* and *Immigration*, d) *Infrastructure* and *European Integration*.

One study type which is facilitated by the subject annotations of parliamentary debates is the investigation of gender differences in the policy areas addressed in the parliament. In Figure 3, we show the distribution of the speakers' gender in relation to speaking time, main subject and number of male or female speakers in DPCv2 (percentage of female and male politicians during the covered time span). The figure confirms on more data a preceding Danish pilot study (Hansen, 2018) indicating that female members of the parliament only speak more than male members in "softer" policy areas addressing *Social Affairs*, *Health* and *Education*. This also confirms the findings in the Swedish parliamentary debates by Paxton et al. (2007). Moreover, our data shows that female members of the Danish parliament in the years covered by DPCv2 have debated as frequently as male members about *Environment*, and nearly as much as them about *Immigration*, *Local and regional affairs*, and *Culture*. The subjects, which are less frequently addressed by female politicians and more often by male politicians are *Infrastructure*, *Foreign Affairs*, and *Economy*. In general, women spoke slightly less than men in the considered period.

5. Supervised Multi-label Classification of Subjects

In this section, we first present related studies and then we describe our classification experiments.

5.1. Related Studies

Supervised text classification algorithms assign predefined classes to text documents when trained on features extracted from the texts. Various features have been used comprising language models over words, lemmas or characters, the number of words and/or lemmas, the sentences' length, bag of word (BOW), term frequency–inverse document frequency (TF*IDF) values and word embeddings (Allahyari et al., 2017; Joulin et al., 2017). The automatic topic classification of policy agendas has been addressed in different studies, e.g. (Quinn et al., 2006; Purpura and Hillard, 2006; Hillard et al., 2008). For example, Hillard et al. (2008) implement a system for classifying the USA's congressional bills. The training data consisted of bills annotated according to the PAP annotation scheme. Good prediction accuracy (between 81% and 88.7%) was achieved using a support vector machine. Karan et al. (2016) test a number of classifiers on a dataset of 7,300 political agendas, manually coded according to the CAP annotation scheme. One of their findings is that a hierarchical approach consisting in first classifying the main subjects, and then their under-subjects gave the best results. The accuracy for the classification of the main subjects reported in the study is 0.77, while the accuracy obtained when classifying the subtopics is 0.68.

Bilbao-Jayo and Almeida (2018) train Convolutional Neural Networks on election manifestos annotated manually by political scientists to classify the political discourse in on-line social networks. The best macro F1-score for the subject classification of the manifestos was 0.75 while that obtained on tweets was 0.63.

The semi-automatic classification of the agenda items in Danish political data has been addressed by Loftis and Mortensen (2018). The authors apply a Naive Bayes classifier to agendas of the Danish city councils' meetings. A part of the agenda corpus was manually annotated using the most common CAP classes. Stemming was applied to the words in the one sentence agendas. The best accuracy results on the data is reported to be between 0.67 and 0.75 depending on the meeting types. The best results were obtained for the larger city councils where the politicians seem to follow the agendas more precisely than in the smaller municipalities.

Hansen et al. (2019) describe classification experiments aimed to identify the main subject in a balanced subset of the first pilot subject annotations of the Danish Parliament corpus. The experiments were run on approx. 19,000 speeches annotated with 18 subject classes. One subject class *Territories* was excluded because it only occurred few times in the pilot data.

Three classifiers were trained on BOW and TF*IDF values of the lemmatised titles of the meeting agendas. The best results (F1-score $= 0.96$) in these experiments were obtained by a multinomial Naive Bayes classifier trained on BOW values. The automatic classification

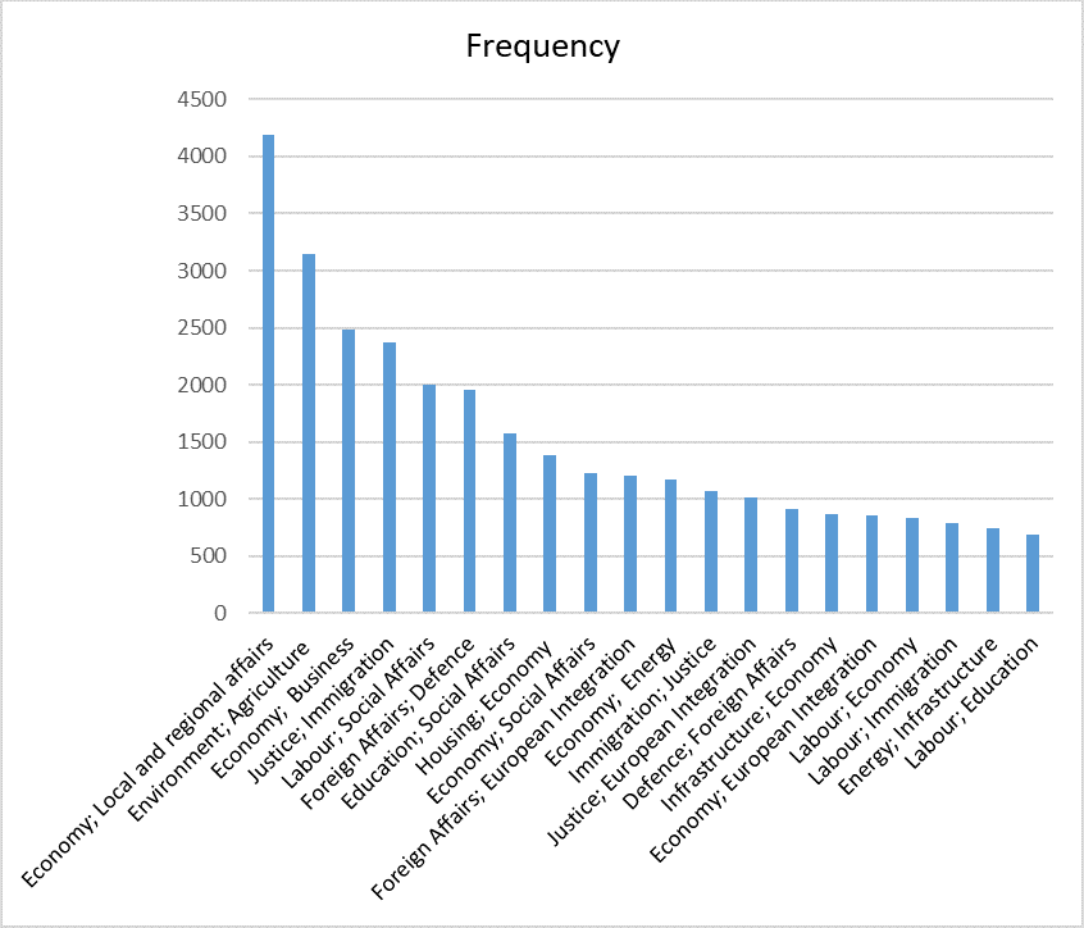

Figure 2: Most frequently co-occurring subject

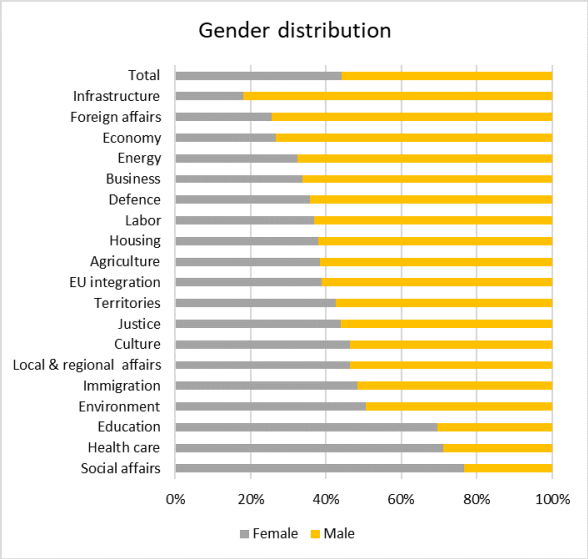

Figure 3: Distribution of speaker's gender w.r.t. speaking time and subject in DPCv2

of the speeches using the BOW values of the speeches' lemmas resulted in an F1-score of 0.74 (Hansen et al., 2019). This result is in the same magnitude of that obtained by Bilbao-Jayo and Almeida (2018) for classifying the main subjects in their manifestos.

5.2. Multi-label Classification Experiments

Differing from the preceding studies, we address the multi-label classification of co-occurring subjects in the released DPCv2, using the words of the agendas or the speeches with and without information about the speakers. The main aims of the classification experiments are the following:

1. Determine whether the annotations of main and secondary subjects in the Danish parliament corpus are consistent and can be reproduced by classifiers.

2. Determine the performance of multi-label classification on three different training sets: a) BOW and TF*IDF values obtained from the words in titles of the agenda meetings, b) BOW and TF*IDF values of the lemmas of the speeches, and c) the data as in the former two items (a and b) enriched with information about the speakers.

3. Investigate the performance of a number of classi-

fiers on these datasets.

Thus, we address supervised multi-label classification aimed to predict the two subjects annotated in the DPCv2. Furthermore, we investigate the impact of information about the speaker on classification, and we test the performance of different classifiers on this task. In the present work, we do not balance the data.

5.3. Pre-processing

We first extracted all speeches annotated with two subjects from DPCv2, then we tokenised, part of speech tagged and lemmatised the transcriptions of the speeches annotated with two subjects. Successively, punctuation marks were removed from the results. These pre-processing steps were made using the CST-tools[7] available from the CLARIN.DK repository[8]. Then, we removed from the data speeches consisting of less than 100 words in order to reduce the amount of data and take into account the observation that some frequent classification errors occur in relation to short speeches (Hansen et al., 2019). Finally, we removed from the data speeches that were classified with very rare combinations of subject 1 and subject 2 (less than three occurrences). The resulting dataset consists of 22,203 speeches annotated with 186 combinations of subject 1 and subject 2.

The data was split in a training set (70% of the data) and a test set (30% of the data). The results of the best classifier in both classification experiments were then validated with 5-fold cross-validation. Since the dataset is imbalanced, stratification sampling was applied ensuring that both training and testing data contain the same percentage of samples for the various subject combinations.

The speaker's features included in the classification experiments are gender, party, role and age.

5.4. Multi-label Classification

Multi-label classification was performed using the python 3 libraries scikit-learn and scikit-multilearn. The multi-label classification module addresses classification as a binary problem (one class against all the others). We used the chain model because it gave the best results. The classifiers in the experiments were multinomial Naive Bayes (NB), support vector machine (SVM) with a linear kernel and a multilayer perceptron (MLP). We applied the MLP with the standard parameters in scikit-learn though extending the number of iterations to 8000 and lowering the number of layers from 100 to 10. The latter change was applied in order to reduce the running time. The results of a majority classifier that balances its results according to the classes' frequency are used as baseline.

We report the results of classification in terms of precision (P), recall (R) and F1-score (F1). The multilabel classification module calculates them as the weighted averages of the corresponding measures for each multilabel. The results of classification from the agenda titles and speaker information are in Table 2. The best

Classif.	Features	P	R	F1
Major.		0.022	0.022	0.022
NB	BOW	0.777	0.956	0.847
	BOW+Speak	0.787	0.726	0.735
	TF*IDF	0.888	0.901	0.884
	TF*IDF+Speak	0.612	0.351	0.415
SVM	BOW	1	0.992	0.996
	BOW+Speak	1	0.995	**0.997**
	TF*IDF	1	0.992	0.996
	TF*IDF+Speak	0.999	0.992	0.996
MLP	BOW	0.988	0.97	0.98
	BOW+Speak	0.954	0.969	0.977
	TF*IDF	0.988	0.985	0.987
	TF*IDF+Speak	0.954	0.94	0.942

Table 2: Results of classifiers trained on the agenda titles and speaker information

results of the classification trained on the agenda titles are obtained by the SVM classifier that achieved an F1-score of 0.997 when trained on BOW values and speaker information. The 5-fold cross validation of the SVM run on the same data (BOW and speaker information) gave an averaged F-score of 0.95, which only represents a fall of 0.047 from the best results obtained in the preceding experiment (Table 2).

All classifiers perform significantly better than the baseline, but the F1-score of NB is worse than that obtained by both SVM and MLP. The results of the latter two classifiers confirm that the agenda titles of the Danish parliament can be used for automatically assigning subjects to the speeches without human intervention. A similar conclusion was made for the annotation of one subject in the classification experiments presented by Hansen et al. (2019).

Looking at the confusion matrices of the SVM on the dataset consisting of BOW values and speaker information, we can see that all classes were recognized with the same high precision and recall. The results of the classifiers trained on features extracted from the speech transcriptions and information about the speakers are in Table 3, which is organised as Table 2. The best performing classifier when applying multi-label classification to predict the two subjects assigned to the Danish parliament's speeches is the MLP trained on the BOW values (F1-score of 0.681). The difference of the results when using BOW or TF*IDF values of the lemmas of the speeches is not large for SVM and MLP, while it is remarkable for the NB whose performance falls dramatically when trained on TF*IDF values with or without speaker information.

The 5-fold cross-validation of the best performing clas-

[7] https://clarin.dk/clarindk/toolchains-wizard.jsp

[8] https://clarin.dk

Classif.	Features	P	R	F1
Major.		0.086	0.0873	0.087
NB	BOW	(0.67	0.609	0.592
	BOW+Speak	0.609	0.733	0.659
	TF*IDF	0.791	0.086	0.145
	TF*IDF+Speak	0.623	0.045	0.078
SVM	BOW	0.652	0.632	0.638
	BOW+Speak	0.651	0.634	0.638
	TF*IDF	0.716	0.656	0.662
	TF*IDF+Speak	0.714	0.656	0.662
MLP	BOW	0.757	0.631	**0.681**
	BOW+Speak	0.751	0.622	0.672
	TF*IDF	0.704	0.666	0.671
	TF*IDF+Speak	0.7	0.662	0.665

Table 3: Results of classifiers trained on speeches and speaker information

sifier, which was the MLP trained on the BOW values of the speeches' lemmas, gave an F1-score of 0.631. Again the fall in performance with respect to the results of the same classifier in the previous experiment is not high (0.05).

6. Discussion of Results

The results of the multi-label classification experiments trained on the agenda titles of the Danish Parliament indicate that the automatic classification based on these titles gives results near to the gold standard represented by the two subjects obtained from the manual annotations of the agenda titles.

All tested algorithms perform significantly better than the baseline, but the best performing algorithm on this task is SVM, and the second best is MLP. Both algorithms perform well independently from the dataset.

In the second group of classification experiments act to identify two subjects from the parliamentary speeches, the best results are obtained by the MLP classifier trained on BOW values of the speeches' lemmas ($F1 = 0.681$). The 5-fold cross-validation shows a small fall with respect to this score confirming the good results. Both SVM and MLP work well with all datasets, while recall for NB is bad when TF*IDF values are used, independently from the presence or not of speaker information. The fact that SVM and MLP work better than NB on the data is not surprising, since NB works best on smaller datasets. The fall in recall when using TF*IDF values is surprising, and we do not have an explanation for it.

In general, the results of multi-label classification are satisfactory given the high number of classes and the fact that the data is skewed. In the future, we should investigate whether some of the erroneous classifications are due to the fact that some of the speeches address three and not only two subjects[9].

The positive results of the second group of experiments are a clear indication that the politicians in the Danish parliament follow the meetings' agendas in the course of the debates. It has been earlier noticed that the correspondence between meetings' agendas and the content of the meetings varies greatly in the case of Danish city councils, whose members do not often have great experience in discussing policy areas (Loftis and Mortensen, 2018). This is certainly not the case for the Danish Parliament where the meetings' agendas are strictly followed during the meetings also with the support of the Speaker who guides the debates.

Looking at the confusion matrices for the various subject classes, we can see that the most frequently occurring subject combinations are always identified correctly, while this is not the case for the more rarely co-occurring subjects.

In the present experiments, we used simple BOW and TF*IDF values of the words of the agendas' titles and the lemmas of the speeches, but word embedding models should be tested in the future. Moreover, we reduced the amount of speeches processed, and we did not investigate the best parameters for MLP, or tested the performance of other neural networks[10].

7. Conclusion and Future work

In this paper, we have presented the Danish Parliament Corpus - with subject annotations, v2. The corpus was recently released in the CLARIN-DK infrastructure.

A first analysis of the subject annotations was described, and multi-label classification experiments act to verify the consistency of the annotations of two subjects in the relevant part of the corpus were accounted for.

We showed the subjects that most often are discussed together in the debates, and we investigated the subjects which are more often addressed by women than by men. Our analysis confirm preceding studies of gender differences in the Swedish and Danish parliament (Paxton et al., 2007; Hansen et al., 2018) indicating that female politicians often address "softer" subjects than male politicians such as *Social Affairs*, *Health care*, and *Education*. Our study also indicates that female members of the Danish parliament talked as much as male members about *Environment* and nearly as much about *Immigration*, *Local and regional affairs* and *Culture*.

In multi-label classification experiments, the agenda titles with or without information about the speakers were used as training data and gave results similar to those obtained by humans extending the manual annotations of the agenda meetings.

[9] In the pilot annotations of the DVCv2 corpus up to three co-occurring subjects were annotated (Hansen et al., 2019). A maximum of two subjects were given the final version of the corpus in order to simplify the data.

[10] Running each series of experiments took already more than 24 hours on a PC with a GPU.

The multi-label classification experiments trained on the speeches' lemmas and information about the speakers also gave good results (F1-score just under 0.7). These results are promising especially considering the high number of subjects' combinations and the large variation in the frequency of the different combinations. All these results indicate that the politicians in the Danish Parliament follow the meetings' agendas during the debates. Therefore, the strategy proposed by political scientists who use agenda titles for annotating the policy areas of political meetings works well for our data.

In the future, we will use the DPCv2 to analyse how different parties have addressed specific subjects over time, and we will extend the subject annotations to more debates. It could also be interesting to investigate whether the subjects annotations of the Danish Parliament debates can be used to annotate policy activities in other types of political data and for other languages. In the classification experiments, we did not investigate the optimal parameters for the MLP classifier, so the results of classification could be improved. Moreover, our corpus can be used for testing other multi-class and multi-label classifiers as well as other feature models.

8. Acknowledgements

This work describe in the paper was done under CLARIN-DK, the Danish part of the European CLARIN infrastructure. CLARIN-DK is funded by the Department of Nordic Studies and Linguistics and the Faculty of Humanities at the University of Copenhagen.

We want also to thank professor Christoffer Green-Pedersen, Department of Political Sciences at the University of Aarhus for his guidance in the choice of the classes to be included in the subject annotation scheme.

9. Bibliographical References

Allahyari, M., Pouriyeh, S. A., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., and Kochut, K. J. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *CoRR*, abs/1707.02919.

Baumgartner, F. R., Jones, B. D., and Wilkerson, J. (2011). Comparative Studies of Policy Dynamics. *Comparative Political Studies*, 44(8):947–972.

Bilbao-Jayo, A. and Almeida, A. (2018). Political discourse classification in social networks using context sensitive convolutional neural networks. In *Proceedings of the Sixth International Workshop on Natural Language Processing for Social Media*, pages 76–85, Melbourne, Australia, July. Association for Computational Linguistics.

Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Pančur, A., Rudolf, M., Kopp, M., Barkarson, S., Steingrímsson, S., Çöltekin, Ç., de Does, J., Depuydt, K., Agnoloni, T., Venturi, G., Pérez, M. C., de Macedo, L. D., Navarretta, C., Luxardo, G., Coole, M., Rayson, P., Morkevičius, V., Krilavičius, T., Darǵis, R., Ring, O., van Heusden, R., Marx, M., and Fišer, D. (2022). The ParlaMint corpora of parliamentary proceedings. *Language Resources and Evaluation*.

Darja Fišer, et al., editors. (2020). *Proceedings of the Second ParlaCLARIN Workshop*, Marseille, France. European Language Resources Association (ELRA).

Darja Fišer, et al., editors. (2018). *Proceedings of LREC 2018 Workshop ParlaCLARIN: Creating and Using Parliamentary Corpora*, Paris, France. European Language Resources Association (ELRA).

Hansen, D., Navarretta, C., and Offersgaard, L. (2018). A Pilot Gender Study of the Danish Parliament Corpus. In *Proceedings of the ParlaClarin workshop at the Eleventh International Conference on Language Resources and Evaluation*, pages 67–72. European Language Resources Association, May.

Hansen, D., Navarretta, C., Offersgaard, L., and Wedekind, J. (2019). Towards the Automatic Classification of Speech Subjects in the Danish Parliament Corpus. In *CEUR Workshop Proceedings*, volume 2364, pages 166–174. CEUR workshop proceedings. https://cst.dk/DHN2019/DHN2019.html.

Hansen, D. (2018). The Danish Parliament Corpus 2009–2017, v1. `http://hdl.handle.net/20.500.12115/8`.

Hillard, D., Purpura, S., and Wilkerson, J. (2008). Computer-Assisted Topic Classification for Mixed-Methods Social Science Research. *Journal of Information Technology & Politics*, 4(4):31–46.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain, April. Association for Computational Linguistics.

Karan, M., Šnajder, J., Širinić, D., and Glavaš, G. (2016). Analysis of Policy Agendas: Lessons Learned from Automatic Topic Classification of Croatian Political Texts. In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 12–21, Berlin, Germany, August. Association for Computational Linguistics.

Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Conference Proceedings: the Tenth Machine Translation Summit*, pages 79–86, Phuket, Thailand. AAMT, AAMT.

Loftis, M. and Mortensen, P. (2018). 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`.

Paxton, P., Kunovich, S., and Hughes, M. M. (2007).

Gender in Politics. *Annual Review of Sociology*, 33(1):263–284.

Purpura, S. and Hillard, D. (2006). Automated Classification of Congressional Legislation. In *Proceedings of the 2006 International Conference on Digital Government Research*, pages 219—-225. Digital Government Society of North America.

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., and Radev, D. R. (2006). An automated method of topic-coding legislative speech over time with application to the 105th-108th U.S. senate. In *In Midwest Political Science Association Meeting*.

Zirn, C., Glavas, G., Nanni, F., Eichorst, J., and Stuckenschmidt, H. (2016). Classifying topics and detecting topic shifts in political manifestos. In *Proceedings of the International Conference on the Advances in Computational Analysis of Political Text (PolText 2016)*, pages 88–93. Dubrovnik, Croatia, July.

Zirn, C. (2014). Analyzing Positions and Topics in Political Discussions of the German Bundestag. In *ACL Student Research Workshop*, pages 26–33. ACL.

10. Language Resource References

Hansen, Dorte Haltrup and Costanza Navarretta. (2021). *The Danish Parliament Corpus 2009–2017, v2 - with subject annotations*. distributed in CLARIN-DK-UCPH Centre Repository `https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/44`.