Multi-aspect Multilingual and Cross-lingual Parliamentary Speech Analysis

Kristian Miok [a,*], Encarnación Hidalgo Tenorio [b], Petya Osenova [c],
Miguel-Ángel Benítez-Castro [d] and Marko Robnik-Šikonja [e]

[a] *ICAM - Advanced Environmental Research Institute, West Unversity of Timisoara, Romania.*
[b] *Department of English and German Studies, Facultad de Filosofía y Letras, Universidad de Granada, Calle Prof. Clavera s/n, 18071, Granada, Spain.*
[c] *Faculty of Slavic Studies, Sofia University St. Kl. Ohridski, Sofia, Bulgaria.*
[d] *Department of English and German Studies, Faculty of Social Sciences and Humanities, University of Zaragoza, Calle Ciudad Escolar, Teruel, Spain.*
[e] *Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia.*

Abstract.

Parliamentary and legislative debate transcripts provide an informative insight into elected politicians' opinions, positions, and policy preferences. They are interesting for political and social sciences as well as linguistics and natural language processing (NLP) research. While exiting research studied individual parliaments, we apply advanced NLP methods to a joint and comparative analysis of six national parliaments (Bulgarian, Czech, French, Slovene, Spanish, and United Kingdom) between 2017 and 2020. We analyze emotions and sentiment in the transcripts from the ParlaMint dataset collection, and assess if the age, gender, and political orientation of speakers can be detected from their speeches. The results show some commonalities and many surprising differences among the analyzed countries.

Keywords: parliamentary debates, natural language processing, deep learning, sentiment analysis, emotion detection, metadata prediction, ParlaMint corpora, cross-lingual analysis

1. Introduction

The development of natural language resources and technologies opens new possibilities for social and political sciences [48]. After decades of analyzing individual political speeches and transcripts, natural language processing (NLP) allows orders of magnitude larger studies. Parliamentary corpora are available for many parliamentary democracies and include draft bills, amendments to bills, adopted legislation, committee reports, and transcripts of floor debates. Processing these heterogeneous records is challenging. However, the recent ParlaMint project has produced unified corpora of parliamentary debates in 17 European parliaments, making them widely accessible [20]. This allows to broaden the scope of analyses from individual countries to joint issues and differences. Modern monolingual and cross-lingual NLP techniques applied to the collected data can provide new insight into the language

---

*Corresponding author. E-mail: kristian.miok@e-uvt.ro.

used, expression of speakers, as well as similarities and differences in emotions and sentiment in different parliaments.

National parliaments have developed their own code of behavior and speech Maurer and Wessels [38]. The comparison between them is difficult. We propose a novel technological approach, combining monolingual and cross-lingual prediction models with machine translation to analyze political, sociological, and linguistic phenomena. Recent research has shown that such approaches are possible for analysis of social media but they have not yet been applied to parliamentary speech corpora. We analyze parliamentary debates from six national parliaments: Bulgarian, Czech, English, French, Slovene, Spanish, and United Kingdom (UK) in the period from 2017 to 2020.

The main contributions of our work can be summarized as follows:

(1) A methodological framework for a comprehensive comparison of parliamentary speeches in cross-lingual setting using cross-domain transfer learning.
(2) Comparison of linguistic effects based on age, gender, and political position of speakers for six parliaments.
(3) Comparison of sentiment and emotions in six parliaments extracted with prediction models.
(4) A novel approach to analyse difference in sentiment based on speakers' age using Bayesian statistical models.

The paper is structured into five further sections. In Section 2, we present background and related work on analyzing parliamentary speech split into age, gender, sentiment and emotions. Section 3 describes the used ParlaMint datasets. In Section 4, we present our methodology, split into prediction of metadata (age, gender, and political wing), sentiment, and emotions. The results are covered in Section 5, while we draw the conclusions and present ideas for further work in Section 6.

2. Background and Related Work

In this section, we first present the necessary background information, discussing also aspects not strictly related to the parliamentary context. Sections 2.1, 2.2, 2.3, and 2.4 cover related works referring to age, gender, sentiment, and emotions in the political discourse. Finally, in Section 2.5, we outline the use of Bayesian methods in parliamentary contexts.

2.1. Language and Age

Age as a factor of language variation is one of the most salient and productive objects of research in the field of sociolinguistics [44]. The description of differences between young and older generations focuses on (in)formality [34, 56]. In the research of group membership through speech Ghafournia [23], sociolinguists describe two types of prestige, overt and covert prestige. Overt prestige is related to standard and more formal linguistic features, which are normally associated with those who hold more power and status. Covert prestige, on the other hand, is the non-standard variety employed in a scenario that encourages cooperation, communality, communication ease, and engagement [57]. Based upon these considerations, some linguistic differences age may explain are adults' preference for syntactic complexity [22], swearing [31], lexical conservativism [32], usage of positive politeness strategies [19], teenagers' tendency towards language change [40], the use of slang [50] or abruptness [13].

Several authors covered the prediction of age and other personal traits such as gender or political affiliation, e.g., Dahllöf [12], who analyzed the wording of political speeches in Swedish. The results show that it is possible to classify politicians according to their age, ideology, and gender to some degree. We analyze six parliaments at once, which opens a broader perspective and gives more general conclusions.

2.2. Language and Gender

Since 1922, a number of studies have addressed the role of gender in language expression – for an overview, see Hidalgo-Tenorio [27]. For example, the debate ranges on whether gender is a social construct, whether there exist different genderlects with different characteristics, and whether a so-called "women's language" is the result of culture or power relations [10], [35]. The linguistic features claimed to characterize females range from articulatory phonetics and grammar to pure pragmatics, e.g., the tendency for hypercorrection, conservativism, self-disclosure and attentiveness; abundance of intensifiers and restricted vocabulary associated with domesticity; preference for simple syntax, minimal responses, emotion(al) language, expressive speech acts, diminutives and terms of endearment; usage of rising intonation, questions and epistemic modality to mark their lack of confidence; and, finally, neither swearing nor turn-taking control, interruption or topic selection in conversation.

In our work, we predict the gender of speakers available as metadata. In this way, we establish a level of differences between speeches used by members of parliament (MPs) of different genders. In the gender detection, we find some interesting research that successfully applies machine learning and/or sentiment analysis [4, 33, 39, 47]. An important consideration in the prediction of speakers' gender is grammatical gender. In the four of the six languages we cover, Bulgarian, Czech, Slovenian, and Spanish, there are three grammatical genders (masculine, feminine, and neuter); in French there are two genders (masculine and feminine), while English has no grammatical gender. Grammatical gender can generally be inferred from the ending of nouns, adjectives, determiners or past participles. In some cases (not all), this means that the gender of a speaker can be determined. Next, we give some examples of these phenomena for the analyzed languages.

BG: Az sam sigurna (I am sure: feminine), Az sam siguren (I am sure: masculine). In Bulgarian, there are synthetic and analytic tenses/moods. The former contains no indication of gender. Here is an example of a synthetic form: Az kazax (I said - feminine and masculine). When an analytic form is used, the gender is indicated by the past participle: Az bix predlozhila (I would suggest: feminine) vs. Az bih predlozhil (I would suggest: masculine).

CZ: Já bych řekl (I would say: masculine), Já bych řekla (I would say: feminine). Again, if a synthetic form is used, then there is no indication of a specific gender: Já si myslím (I think: feminine and masculine).

ES: Estoy harta del populismo (I'm sick of populism: feminine); Estoy harto del populismo (I'm sick of populism: masculine).

FR: Je suis prête (I am ready: feminine), Je suis prêt (I am ready: masculine). Again, there are many cases where the gender is not revealed, e.g., Comme j'ai dit (As I said: feminine and masculine).

SI: The gender is revealed when using the first person singular in the past and future tense, e.g., Rekla sem (I said: feminine), Rekel sem (I said: masculine). The gender is not revealed in the present tense, e.g., Mislim (I think: both masculine and feminine).

In English, the gender is not a grammatical category but a lexico-sematic feature that can be inferred from the personal and relative pronouns used (the person who arrived; he is nice) and a few morphemes (actor vs actress; policeman vs policewoman); adjectives, determiners or past participles do not show it (this happy man vs this happy woman; she was kissed vs he was kissed). These features are not revealing of the speakers' gender.

2.3. Sentiment in Politics

The problem of computational sentiment analysis for parliament discourses has been tackled extensively but with relatively little cross-country comparison. In most cases, sentiment analysis involves document, sentence, and aspect-level analysis.

The research of Dzieciatko [17] and Rheault et al. [48] have applied sentiment analysis to entire corpora at the highest granularity, i.e. their analysis of the Polish and UK parliaments aggregates sentiment scores of all speeches. Honkela et al. [28] explore the overall sentiment of EU Parliament transcripts on the dataset level, whereas Sakamoto and Takikawa [53] consider the polarity of US and Japanese datasets.

While in NLP sentiment analysis is often fine-grained (such as at the level of speech, speech segment, paragraph, sentence, or phrase), in political science, the unit of analysis is primarily an actor (individual politician whose contributions are pooled together). This is the focus of most works on position scaling, a task very much associated with that field. It appears that this confirms to some extent Hopkins and King [29] assertion that while computer scientists are interested in finding the needle in the haystack, social scientists are more interested in characterizing it. The exceptions come from works in the social and political sciences [29, 30] that propose ways to optimize speech-level classification for social science purposes and from computer science [24], which also consider the position scaling issue.

Sentiment detection has advanced considerably in the last few years with the advent of large pretrained language models such as BERT [16]. This has allowed applications to social media, stock market predictions, user stance detection in reviews, hate-speech detection, etc. However, parliamentary discourse is hard to analyze for established techniques due to specific formal speech and linguistic differences to existing training datasets [48]. Rudkowsky et al. [51] study several machine learning approaches based on word embeddings for Austrian parliamentary speeches. Similarly, Abercrombie and Batista-Navarro [1] and Elkink and Farrell [18] investigate predicting votes based on the parliament speeches.

In this paper, we follow political sciences and predict the sentiments of speakers based on their speeches. The results are cross-lingual for six different parliaments, which, to our knowledge, has not been done before.

2.4. Emotions in Politics

Alba-Juez [2] states that whatever we say, write, hear, and read is produced and processed through the filter of affect. Cognition and emotion are, therefore, two mutually interconnected systems [6]. In this regard, Van Dijk [58] argues that one of the most distinguishing features of manipulation lies in shaping and framing messages in such a way that they accord with their recipients' negative emotions, usually deriving from feelings of powerlessness and injustice. In the current political landscape, which is imbued with populism, this idea is of utmost importance, especially at a moment when the emotional is preferred to the intellectual.

Research shows how resentment, anxiety, panic, anger, and disgust can help populist politicians seduce their voters [7]. They may use the same discursive strategies to attack and bring their rivals into disrepute. For instance, they can spread unreliable news about their opponents and other sensationalist information with bombastic but simple expressions; plentiful negative ethical and aesthetic evaluative terms; swear words and colloquialisms; and adversarial vocabulary echoing 20th-century propaganda[1]. Despite the similarities, however, the discourses of right- and left-wing populist leaders are quite diverse. Open opposition to capitalist elites drives left-wing populists to show their hatred of big corporations, financial, and governmental institutions [14]. On the other hand, due to their fear of losing their status because of the alleged privileges granted to minority groups in a multicultural society, right-wing populists cannot conceal their antagonism and hostility towards such communities [54].

Our analysis is unique in detecting and comparing emotions in six national parliaments at once. This reveals some similarities but also surprising differences.

2.5. Bayesian Methods for Parliamentary Data

Bayesian methods allow for the estimation of the posterior probability of a model given data, which provides a measure of both the quality of the model and the uncertainty of the estimates. This is particularly useful when dealing with small data sets or noisy data. In our previous work, we applied Monte Carlo Dropout to deal with prediction uncertainty [41, 42] in natural language processing. Within the parliamentary context, Hansen [26] estimated the positions of the Irish parliamentary parties using the Bayesian ideal point estimation framework and showed that it is possible to distinguish only two blocs of parties in each period, one block supporting the government and one forming the opposition. Han [25] explored the performance of spatial voting models to the roll calls of the EU parliament. Montalvo et al. [43] proposed a new methodology for predicting electoral results that combines a fundamental model and national polls within an evidence synthesis framework.

In this work, we propose to use the Bayesian AB testing to determine the age threshold with the largest difference between positive and negativce sentiment. In this way, we avoid setting a fixed threshold for all parliaments and get interesting differences between the tested parliaments.

3. The Data

In this section, we describe the datasets used in our analysis. Section 3.1 describes the ParlaMint project, which collected and preprocessed the data, while in Section 3.2 we provide information on the actually used datasets and graphically present distributions of MP's age and gender across parliaments.

3.1. ParlaMint Project Background

ParlaMint[2] project aims to enhance the development and usage of national parliamentary corpora. The data has been synchronized with respect to the same TEI format and time span. It can be exploited for linguistic, social, and political research in cross-lingual and cross-parliament settings.

---

[1] https://www.thebritishacademy.ac.uk/blog/how-language-fake-news-echoes-20th-century-propaganda/
[2] http://www.clarin.eu/parlamint

We use the multilingual comparable corpora of parliamentary debates ParlaMint 2.1 containing parliamentary debates mostly starting in 2015 and extending to mid-2020, with each national corpus containing various amounts of words varying from $800\,000$ words (for Hungarian) to $109\,000\,000$ words (for UK). The sessions in the corpora are marked as belonging to the COVID-19 period (after November 1st 2019), or being "reference" (before that date). The data is freely available through the CLARIN.SI repository[3].

The corpora contain extensive metadata, including many aspects of the speakers (name, gender, MP status, party affiliation, party coalition/opposition). The data are structured into time-stamped terms, sessions, and meetings. Speeches are marked by the speakers and their roles (e.g., the chair or regular speaker). The speeches also contain marked-up transcriber comments, such as gaps in the transcription, interruptions, applause, etc. More information about the creation of the corpora, the common standard, and specifics of each national corpus can be found in Erjavec et al. [20].

3.2. The Datasets

At the time of writing this paper, the ParlaMint project released data for 16 languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, French, Hungarian, Icelandic, Italian, Latvian, Lithuanian, Polish, Slovenian, Spanish, and Turkish. In total there are $3\,774\,204$ utterances and $494\,949\,904$ words. The quality of the textual corpora and metadata varies across the languages. In our experiments, we studied parliaments in six countries: Bulgaria (BG), Czech Republic (CZ), France (FR), Slovenia (SI), Spain (ES), and the United Kingdom (UK). The criteria for this selection were mainly the quality of the provided corpora and that we, as authors, understand the languages and the political situation in these counties. The available data for the specific year varies across the parliaments, as shown in Table 1. We decided to analyze data from 2017 to 2020 expecting this selection to be the most informative and provide the most interesting insights.

The data per parliament and year are organized as the parliament session documents. Every text document with the talks is paired with the document containing the session metadata such as title, time, term, number of session and meeting. This supplement document also includes the speaker information (speaker type, speaker party, party parliament status, speaker name, speaker gender, and speaker birth). The number of parliament session documents per country and year are presented in Table 2. The number of sessions varies per parliament. For the Czech parliament, the number is the largest, while Slovene and Spanish parliament exhibit lower numbers of sessions.

Table 1

Number of words per parliament and per year in the analyzed part of the ParlaMint corpus.

Year	BG	CZ	ES	FR	SI	UK	Σ
2017	3,052,523	2,959,956	2,553,736	5,584,411	4,489,429	16,715,363	35,355,418
2018	4,342,676	3,487,291	3,175,000	13,470,543	2,842,075	19,675,067	46,992,652
2019	3,349,615	4,003,544	885,134	11,864,705	3,601,000	16,937,349	40,641,347
2020	2,240,914	4,299,581	3,697,803	5,767,864	1,885,329	19,435,104	37,326,595
Σ	12,985,728	14,750,372	10,311,673	36,687,523	12,817,833	72,762,8834	160,316,012

---

[3]http://hdl.handle.net/11356/1432

Table 2

Number of the sessions per parliament and per year in the analyzed part of the ParliaMint corpus.

Year	BG	CZ	ES	FR	SI	UK	Σ
2017	101	711	64	156	74	259	1365
2018	132	635	71	362	65	309	1574
2019	67	784	18	331	62	281	1543
2020	75	710	65	171	33	313	1367
Σ	375	2840	218	1020	234	1162	5849

We show the distribution of the number of speeches based on MP's age and gender across different parliaments in Figures 1 and 2, respectively. To get representative age cohorts, we merge MPs into 5 year intervals for 1. The distributions based on speakers' age tend to peak between 50 and 60, except for France where there are no peaks. Concerning gender, there are considerably more speeches of male MPs in all parliaments, with the smallest difference in gender in Spain and the largest in Slovenia and Czechia.

Bulgaria.

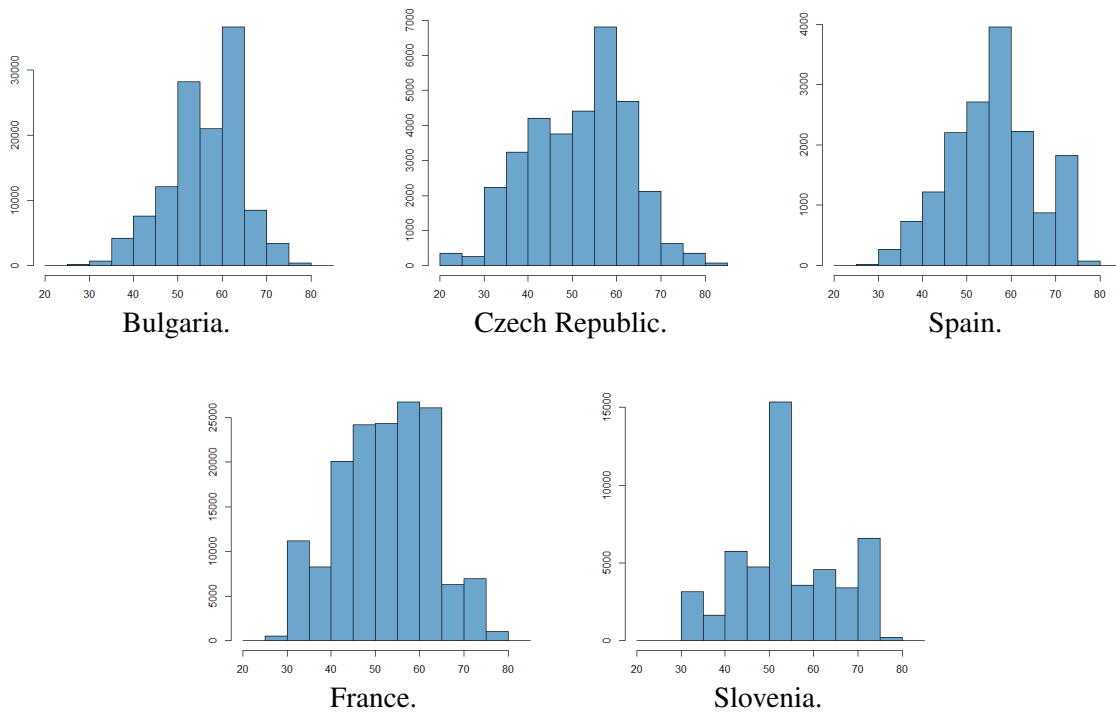Czech Republic.

Spain.

France.

Slovenia.

Fig. 1. The distribution of the number of speeches relative to the MP's age across the parliaments. The age information is not available for the UK.
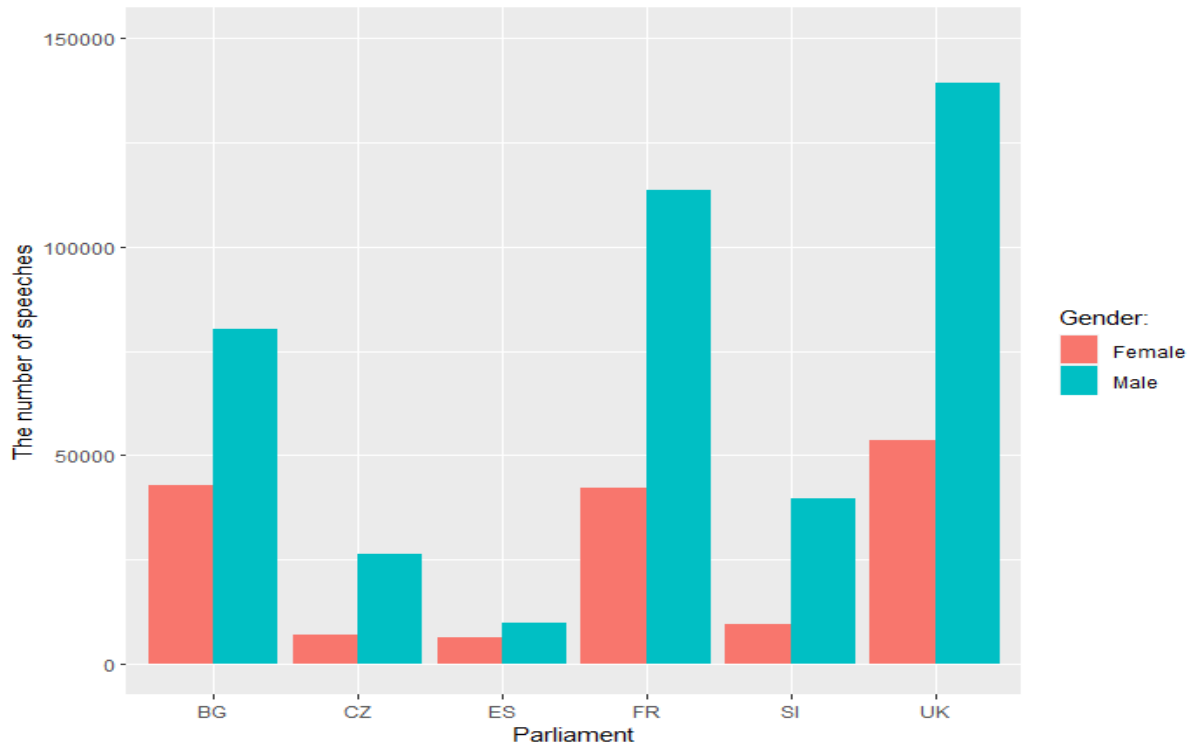
Fig. 2. The distribution of the number of speeches relative to the MP's gender across the parliaments.

4. Methodology

We selected several modern NLP approaches to analyze the speeches in chosen languages. We use supervised text classification for sentiment and emotion analysis, as well as to predict the age, gender, and political wing of a speaker. Since 2019, a standard approach to text classification has been fine-tuning one of the large pretrained language models such as BERT [16] to the specific task. We followed this approach and used multilingual BERT (pretrained on 104 languages) to predict desired variables in a uniform way accross the tackled languages. The trained models were used to predict speakers' meta-information (gender, age, political wing), as described in Section 4.1, as well as their sentiments and emotions based on the language and contextual information in speeches, as outlined in Sections 4.2 and 4.3, respectively. The details are presented below.

4.1. Meta Data Prediction

Datasets from the ParlaMint project contain information about the speakers' age, gender, and political party. Thus, we fine-tune the multilingual BERT model to predict each of the three metadata variables from individual speeches of parliament members. The prediction accuracy of the models reveals the amount of information about the metadata stored in the parliament speeches. The variables that we predict are:

(1) Age of the speaker. The original corpora contained the birth year of the speaker, from which we computed the speaker's age. We dichotomized the age into two groups, separately for each country, using the Bayesaian AB testing to maximize the difference in sentiment between the younger and older group of MPs. The details of this process are contained in Section 5.3.
(2) Gender of the speaker. The original datasets contain a meta-data variable with the value 'F' for female and 'M' for male speakers, which we converted into 0 for females and 1 for males. We randomly selected 2500 speeches by male and 2500 speeches by female MPs.
(3) Political wing of the speaker. We assumed that both left and right political parties can be split into moderate (center-leaning) and extreme (far-left and far-right). We separately predicted the differences between the moderate parties (i.e. center-left and center-right) and extreme parties (extreme-left and extreme-right). For both center-leaning and extreme party comparisons, we selected 2500 speeches from the left and 2500 speeches from the right-wing of the political spectrum.

Each of the created datasets was split into training and testing data parts (80 % for training and 20 % for testing). The training set was used to fine-tune the multilingual BERT (mBERT) model[4]. BERT [16] is a text representation model based on the transformer neural network architecture [59], pretrained on the masked language modeling task using a large corpus of data. The mBERT is pretrained on 104 languages. In our work, we used the pretrained models available in the HuggingFace platform and fine-tuned them separately for each task.

4.2. Sentiment Prediction

For automatic classification of sentiment in text data, various approaches have been developed, the most successful being machine learning classifiers trained on human-annotated corpora. The main challenge of these approaches is that they tend to be domain-specific and work best when trained with labeled data from the target domain but are less effective in other domains. However, as producing labeled datasets is expensive, researchers often apply the trained models across domains and languages. The cross-lingual transfer is possible either by the machine translation from a language without a suitable dataset to a language where such a dataset exists or by using a pretrained multilingual language model such as mBERT.

As there are no specific parliamentary language sentiment datasets, our cross-lingual and cross-domain approach relies on a collection of sentiment datasets from different languages in the domain of news and media. The reason to choose the news sentiment datasets is that the language and the context used are relatively similar to the parliament discourse. We use two-class sentiment prediction with the negative sentiment labelled 0 and the positive labelled 1. As our previous work has shown that sentiment classifiers can be successfully transferred between languages (especially similar ones) with mBERT [49], we combined datasets from three languages as follows.

(1) *Slovenian SentiNews dataset* [8] is a manually labeled sentiment dataset containing 10 427 documents there were annotated on a document, paragraph and sentence level. For our task, we used the document level annotation selecting the negative and positive labeled news. The selected instances consist of 3337 negative and 1665 positive Slovene news.
(2) *English news headlines datasets*, consisting of two sources:

---

[4]https://huggingface.co/bert-base-multilingual-cased

- The financial news headlines dataset [37] was labelled with the sentiment from the perspective of a retail investor and constructed based on the human-annotated finance phrase bank. The data contained 604 negative and 1363 positive headlines.
- The SEN dataset [5] is a recent human-labelled dataset for entity-level sentiment analysis of political news headlines. The dataset consists of 3819 human-labelled political news headlines from several major online media outlets in English and Polish. Each record contains a news headline, a named entity mentioned in the headline, and a human-annotated label (positive, neutral, or negative). The original SEN dataset package consists of two parts: SEN-en (English headlines that split into SEN-en-R and SEN-en-AMT), and SEN-pl (Polish headlines). The English dataset names are comming from the way the annotation process was done. For SEN-en-R each headline-entity pair was annotated via the open-source annotation tool doccano[5] by at least 3 volunteer researchers while for the SEN-en-AMT the Amazon Mechanical Turk service was used. For our task, we selected only the labeled instances from the two English datasets that were annotated as negative and positive, ending with 982 negative and 456 positive instances.

(3) *Russian news dataset* obtained from the Kaggle[6] that contains 8000 sentiments annotated news in the Russian language. From these, we selected 1434 negative and 2795 positively labeled instances.

By combining all the above datasets, we obtained our final training dataset with 12 636 labeled instances, of which there are 6357 negative and 6279 positive.

4.3. Emotion Detection

Besides informative content, texts also communicate attitudinal information, including emotional states [3]. As there exist many emotional states and feelings, the emotion detection is a much more challenging compared to the sentiment analysis, where typically three attitudes (positive, negative or neutral) are predicted from the given user input. In contrast to that, the emotion detection task deals with both primary emotions (e.g., happiness, sadness, anger, disgust, fear and surprise) as well as more complex emotion models involving different dimensions of emotions and psychology theories [52]. Apart from this, emotion analysis from texts suffers from relatively small and homogeneous annotated corpora [11]. While there exist several English language datasets, emotion analysis for less-resourced European languages is much more problematic [45]. Most languages have very few, if any, well-annotated dataset that can be used to train text classification models. To overcome this issue, we use multi-lingual and cross-lingual approaches and restrict the covered emotions to positive and negative to get better statistical coverage and also reduce the error due to machine translations and/or cross-lingual transfer.

Similarly to the sentiment analysis, we use the mBERT model to detect emotions in the parliamentary speech. Our preliminary investigation showed that precise detection of many emotions is not possible in the multilingual setting, so we only categorized emotions into positive and negative. Again, the emotion detection datasets for parliamentary domain and our languages (except English) do not exist, so we use several English language datasets from various domains to enable better generalization across domains. We fine-tune the mBERT model with the following four emotion-labelled datasets.

---

[5]https://github.com/doccano/doccano
[6]https://www.kaggle.com/competitions/sentiment-analysis-in-russian/data

(1) The Kaggle Twitter dataset[7] contains 13 different emotions and $40\,000$ records. We selected 5209 happiness, 3842 love, 1323 hate, and 110 anger tweets. The instance were grouped into negative emotions (hate and anger with in total 1433 instances) and positive emotions (happiness and love with in total 9051 instances).

(2) The HuggingFace[8] Twitter dataset [55] contains $16\,000$ annotated tweets. From these, we selected 1937 fear (labeled as negative) and 1304 love instances (labeled as positive).

(3) GoEmotions[9] dataset [15] is a human-annotated dataset of 58k Reddit comments extracted from popular English-language subreddits and labeled with 27 emotion categories. Some comments have multiple emotion labels but we selected only instances with a single labeled emotion. Extracted negative emotions are anger (1025 instances) and disgust (498 instances), while positive emotions are love (1427 instances) and optimism (861 instances). In total, we extracted 1523 negative and 2288 positive instances from this dataset.

(4) XE[10] emotion dataset [46] contains $25\,000$ human-annotated Finnish and $30\,000$ English sentences. From the English dataset, we selected anger and disgust sentences (in total 3803 instances) as negative emotions, and 1721 joy sentences as positive emotions.

Our final emotion detection dataset contains $23\,282$ instances from which 8918 are labeled as containing negative and $14\,364$ as expressing positive emotions.

5. Results

In this section, we report and interpret the obtained results. We present results related to prediction of metadata (age, gender, and political wing) in Section 5.1, and sentiment and emotions analysis in Section 5.2.

5.1. Metadata Prediction: Age, Gender, and Political Wing

As we described in Section 4.1, we find-tuned the multilingual BERT language model to predict speaker's metadata such as age, gender, and political position. The mBERT model was fine-tuned for each of the metadata variables and six countries separately, and we present the predictive performance measured on the testing datasets. Being able to predict any of these three variables indicates considerable differences in the language used by specific groups of parliamentary speakers. The differences and similarities between different countries are discussed below.

5.1.1. Predicting the age of speakers

For each of the analysed parliaments, we split the speeches into three groups according to the first and the third age quartile. For each country, we created two prediction models, one trying to distinguish between the speeches of MPs younger than the first quartile and aged between the first and third quartile (Table 3), and the second distinguishing between the speakers aged between the first and third quartile and speakers older than the third quartile (Table 4). In this way, we investigate language differences between three generations of MPs, checking if their language is age-specific.

---

[7]https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text
[8]https://huggingface.co/datasets/emotion
[9]https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html
[10]https://github.com/Helsinki-NLP/XED

Tables 3 and 4 show that age is a relatively well-predicted characteristic of speakers in Spain, Bulgaria, Slovenia and Czech Republic, while in France, there are very few language differences between speakers of different ages. The higher the prediction performance, the easier it is to distinguish between speakers' age groups, and the language generation gap is more significant.

The only notable difference between Tables 3 and 4 are the scores for the Czech Republik. Here the gap between the youngest and middle-aged MPs is considerably larger than between the middle-aged and older MPs. We hypothesize that this might be the result of transition between the communist-rule and parliamentary democracy, where the language of younger generations is less affected by the previous social system.

Table 3

Classification accuracy, precision, recall, and F_1-score for predicting speakers' age (younger than the first quartile or between the first and third quartile) in different parliaments. For the UK parliament, the age of speakers is not available.

	BG	**CZ**	**ES**	**FR**	**SI**	**UK**
Accuracy	0.71	0.70	0.72	0.53	0.65	/
Precision	0.76	0.71	0.74	0.56	0.66	/
Recall	0.64	0.69	0.77	0.43	0.65	/
F_1	0.69	0.70	0.76	0.49	0.66	/

Table 4

Classification accuracy, precision, recall, and F_1-score for predicting speakers' age (between the first and third quartile or older than the third quartile).

	BG	**CZ**	**ES**	**FR**	**SI**	**UK**
Accuracy	0.66	0.63	0.74	0.52	0.61	/
Precision	0.67	0.64	0.77	0.53	0.63	/
Recall	0.65	0.61	0.75	0.54	0.61	/
F_1	0.66	0.63	0.76	0.53	0.62	/

Below we try to explain the two extreme cases, Spain with the largest gap between age groups and France with the smallest.

Flaherty [21] presents a historical development of French political discourse, which is directed toward uniformity in discursive strategies and may explain their similarities. A similar conclusion was drawn by Lehti and Laippala [36] who show that the language used in French politicians' blogs is relatively standard.

The Spanish case, with the largest differences between younger and older parliamentary speakers, may be explained by the fact that after the end of the two-party system, new parties with younger leaders wanted to contrast with more senior and more socially privileged individuals [9].

5.1.2. Predicting the gender of speakers

As discussed in Section 2.2, the information about speakers' gender may be detected from the grammatical structures used in their speech for all the analyzed languages but English if speakers use phrases related to their personal beliefs and feelings. Another possibility to detect gender is if speakers of different gender indeed use different language. We test differences in the speech between genders in two ways: first, we predict the gender in the original language, and second, we translate all speeches to English,

and predict the gender of the translated speeches, thereby avoiding possible leakage from grammatical structures in the original languages.

As Table 5 shows, gender in the original texts is detectable to some degree in all analyzed countries. Slavic language (Slovenian, Czech, and Bulgarian) speakers express their gender the most explicitly, followed by Spanish, English and French speakers. The last two (English and French) are surprising for different reasons. In French, where gender may be expressed with the language, there is little evidence that speakers express it. Similarly to the age, we hypothesize that the case of French could be explained by the tendency toward language uniformity in French political discourse. Contrary to that, in English, where gender expression is not part of the grammar, the speakers' gender can be detected nevertheless, indicating differences in expression between male and female MPs.

Table 5

Predictive performance of speakers' gender prediction for six *original* parliamentary datasets.

	BG	CZ	ES	FR	SI	UK
Accuracy	0.70	0.86	0.66	0.58	0.88	0.58
Precision	0.69	0.87	0.66	0.56	0.92	0.59
Recall	0.72	0.85	0.71	0.60	0.93	0.64
F_1	0.71	0.86	0.68	0.58	0.93	0.61

In Table 6 we see that gender differences are detectable also if the speeches are first translated to English (as gender neutral concerning grammar), indicating differences in the expression between male and female MPs. For Bulgarian, Spanish, and French, we can observe small differences in prediction scores compared to the original languages (in Table 5), while Czech and Slovene show substantially lower prediction performance for the translated speeches. Nevertheless, in Slovene the gender differences are still the most pronounced.

Based on both original and translated speeches, we can conclude that gender differences exist in all analyzed countries.

Table 6

Predictive performance of speakers' gender prediction for five parliamentary datasets *translated into English*. For UK the translation makes no sense.

	BG	CZ	ES	FR	SI	UK
Accuracy	0.73	0.65	0.66	0.58	0.77	-
Precision	0.74	0.67	0.68	0.56	0.77	-
Recall	0.74	0.62	0.65	0.54	0.79	-
F_1	0.74	0.64	0.67	0.55	0.78	-

5.1.3. Predicting the political orientation of speakers

This section investigates the speech differences between parliament members with different political orientations. Our approach is again based on prediction models that predict the metadata (party membership) available for speakers. A successful prediction would testify that speakers of different political orientations use different language, while low success in prediction would indicate that the compared parties use similar discourse. We investigate two scenarios of different difficulty:

(1) Predicting the left/right positioning of speakers from firmly or extreme left and right political parties. This problem shall not be very difficult, as we expect significant differences in the political stance between these parties, which we assume will be expressed in different content and possibly other linguistic features. The results are presented in Table 7.

(2) Predicting the left/right positioning of speakers from the center-left and center-right political parties. This shall be a more complex problem as we try to distinguish between speakers from relatively similar parties. The results are presented in Table 8.

Table 7

Classification accuracy, precision, recall and F_1 score when predicting the political wing of speakers from the extreme left and extreme right parties in different parliaments.

	BG	CZ	ES	FR	SI	UK
Acc	0.80	0.88	0.80	0.74	0.84	0.78
Pre	0.78	0.87	0.82	0.75	0.88	0.81
Rec	0.77	0.86	0.77	0.70	0.80	0.75
F1	0.78	0.86	0.80	0.73	0.84	0.78

Table 8

Classification accuracy, precision, recall and F_1 score when predicting the political wing of speakers from center left and center right parties in different parliaments.

	BG	CZ	ES	FR	SI	UK
Acc	0.78	0.78	0.70	0.58	0.84	0.76
Pre	0.79	0.8	0.69	0.62	0.86	0.78
Rec	0.78	0.73	0.79	0.48	0.89	0.74
F1	0.79	0.76	0.74	0.54	0.87	0.76

As expected, the differences in speech between extreme left- and right-wing parties are relatively well predictable for all countries, indicating big differences in the discourse of these parties. The classification accuracy between countries ranges from 88% (Czech Republic) to 74% (France).

Surprisingly, the differences are still large between center-left and center-right parties, ranging from 87% (Slovenia) to 54% (France). France is an exception with its low predictability (again, likely due to the tendency for uniform political discourse), which is much more prominent in other countries. For two countries, Slovenia and Bulgaria, the differences between central parties are larger than between extreme parties, which may indicate strong political competition between the central parties.

5.2. Sentiment and Emotions Detection

This section presents the results obtained from the sentiments and emotion detection experiments. For these experiments, we fine-tuned multilingual BERT on the training datasets described in Section 4.2 (sentiment) and Section 4.3 (emotions). First, we try to establish the quality of the trained models. For that purpose, during the fine-tuning process, a small part of the training data (10 % of all training instances) was used for the validation after each training epoch. This classification accuracy is shown in Table 9 for both the sentiment and emotion detection tasks. The results show that sentiment and emotions can be relatively well-predicted, which is a positive indication of the reported results' reliability. Slightly

better results in predicting positive and negative emotions are expected, as emotion datasets are all in English, all collected from social media, and therefore relatively homogeneous. The sentiment detection datasets are multilingual and collected in different domains; thus, the training accuracy reported in Table 9 is expected to be lower. However, this does not mean that the obtained sentiment model would provide less good generalization on our out-of-domain parliamentary data.

Table 9

Classification performance on the validation sets during fine-tuning the mBERT model on sentiment and emotion prediction tasks.

Epoch	Sentiment	Emotions
1	0.85	0.90
2	0.86	0.91
3	0.87	0.92
4	0.87	0.92

To further assess the quality of the produced sentiment prediction model, we selected 20 talks with the highest probability of the negative sentiment for each of the parliaments and manually validated weather predictions are correct. The results are presented in Table 10. Based on the results, we consider the model's accuracy good enough (and comparable to other sentiment prediction models in the literature) to provide a reliable picture of the sentiment in our study. The fine-tuned models were used on our six parliamentary speech datasets. To obtain reliable and comparable statistics, we randomly selected 10 000 speeches that have more than 30 characters of regular parliament members from 2020 for each of the six parliaments. For each speech, the trained mBERT model returned the sentiment score between 0 and 1 (0 indicating the negative and 1 indicating the positive sentiment).

Table 10

Manually determined percentage of instances with correctly predicted negative sentiment in parliamentary speeches for compared parliaments.

Parliament	Accuracy
BG	85 %
CZ	80 %
ES	95 %
FR	90 %
SI	95 %
UK	95 %

Predictions are summarized in Figure 3 and Table 11 from which we can compare the parliamentary sentiment across the countries. Figure 3 shows the histogram of sentiment distribution in each country. The Czech, Spanish and United Kingdom parliaments seem to express less negative sentiment than positive; in the Bulgarian and French parliaments, there seem to be a relatively balanced situation, while the Slovenian parliament shows the least positive sentiment. We attribute the results for Slovenia to the poisonous exchanges between the pro-government and opposition parties at the observed time when the previous opposition took over the government in the middle of the mandate. To get a numeric overview of the sentiment, we set the decision threshold for negative sentiment at 0.2 and for positive sentiment at 0.8 and counted the number of negative and positive speeches. The results are presented in Table 11. As before, we conclude that the parliament with the highest percentage of negative sentiment is Slovenian, while the UK parliament speeches contain the highest positive sentiment rate.
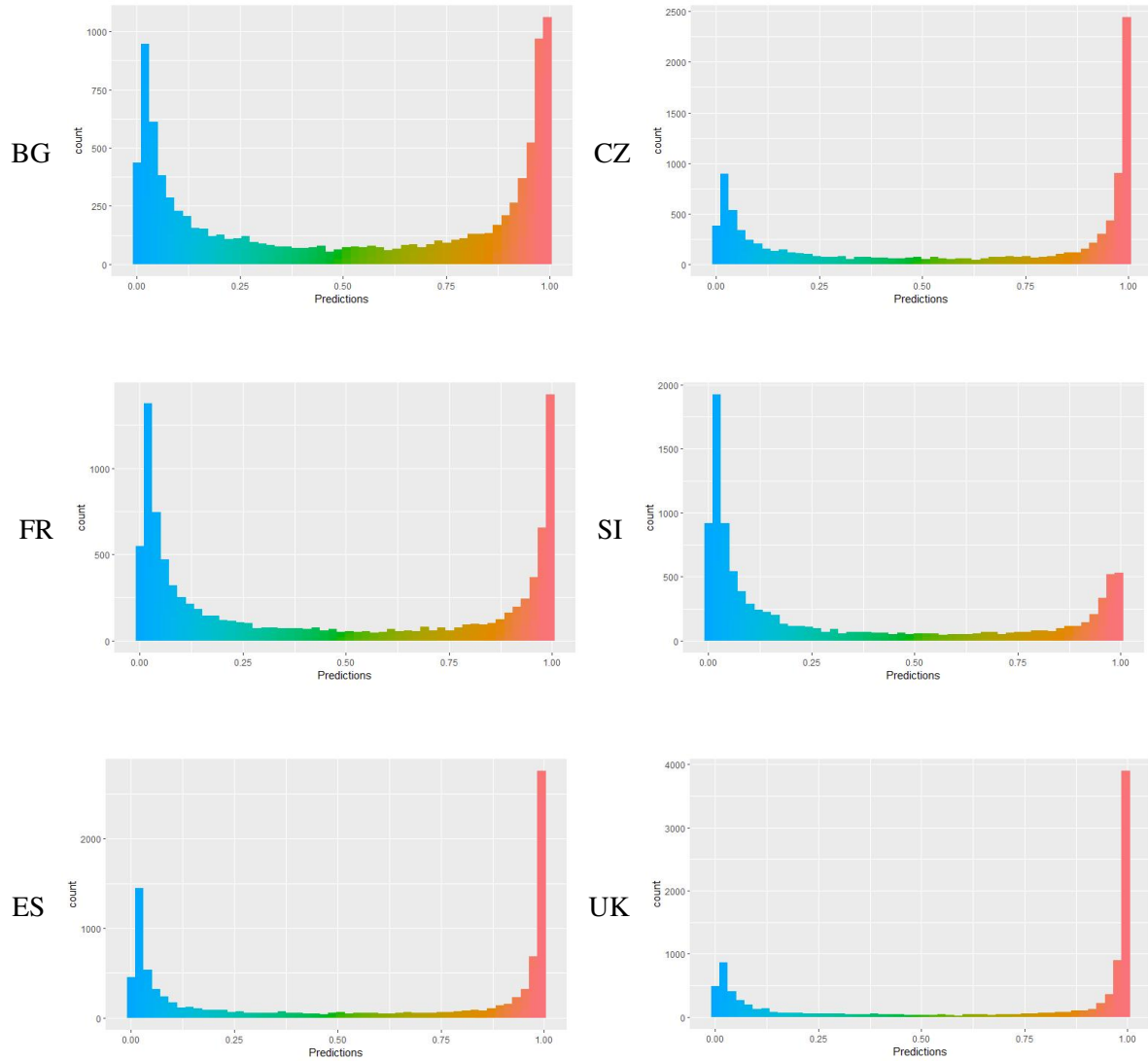
Fig. 3. The histograms showing the distribution of sentiment predictions for six parliaments. The score of 0 indicates completely negative and the score of 1 completely positive sentiment.

Similarly to sentiment, we process emotions. To validate how good the emotions detection models are, we selected 20 speeches predicted to be the most negative for each parliament and manually checked if predictions were correct for them. The results are presented in Table 12. We can observe significantly lower accuracy in all countries compared to the sentiment (shown in Table 10). While this is not surprising as the emotion prediction is considered harder compared to sentiment, this makes the results and interpretations presented below less reliable compared to the sentiment.

We show the results for the emotion detection in Figure 4 (distribution of sentiment predictions) and Table 13 (the percentage of positive and negative emotions, taking 0.2 and 0.8 as the decision threshold

Table 11

Percentage of negative and positive sentiment in parliamentary speeches for observed countries.

Parliament	Negative sentiment	Positive sentiment
BG	35.67	39.68
CZ	31.66	48.58
FR	44.34	34.75
SI	**58.06**	22.14
ES	36.32	46.41
UK	27.34	**59.52**

Table 12

Manually determined percentage of instances with correctly predicted negative emotions in the parliamentary speeches of compared parliaments.

Parliament	Accuracy
BG	45 %
CZ	40 %
ES	65 %
FR	55 %
SI	55%
UK	50 %

values). As the results show, positive emotions are strongly dominant in all countries except France and UK, where positive and negative emotions are almost balanced.

Table 13

Percentage of negative and positive emotions in parliamentary speeches for observed countries.

Parliament	Negative emotions	Positive emotions
BG	16.5	64.86
CZ	14.75	65.27
FR	37.6	41.18
SI	9.71	72.16
ES	20.24	62.31
UK	41.18	44.03

5.3. The Impact of Age on Sentiment

In this section, we investigate the impact of age on the sentiment. We apply Bayesian statistics to find the age which best distinguishes between the positive and negative sentiments of the speakers.

We start with the hypothesis that older speakers more openly express the negative sentiments compared to younger speakers (the reverse hypothesis would be equally suitable for our approach), and form two hypotheses to use in the Bayesian hypothesis test:

H_0: Younger MPs express the same amount of negative sentiment as the older ones.
H_1: Younger MPs express more positive sentiments as the older ones.

To determine the age which best separates the younger from the older MPs in terms of positive sentiments, we estimated the posterior distribution for multiple age cutoff points as shown in Table 14. The resulting numbers were constructed as follows:
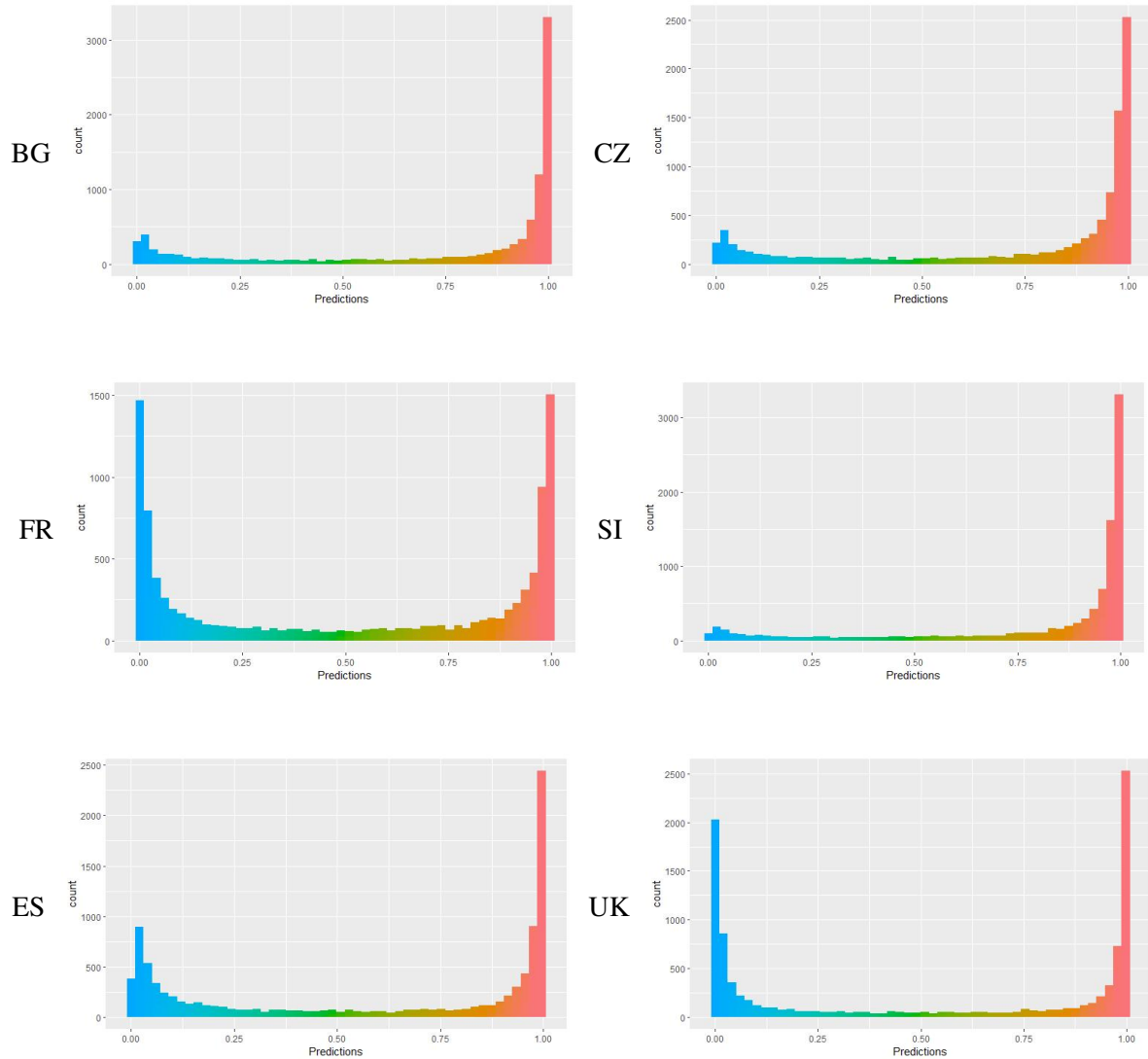
Fig. 4. The distributions of emotion predictions for six parliaments. The score of 0 indicates completely negative emotions and the score of 1 completely positive emotions.

(1) We separated the speakers to the younger and older based on the *Age Cutoff*.

(2) For both, the younger and older population, we dichotomized the sentiment scores to 1 (scores higher than 0.5) or 0 (scores lower than 0.5), and assumed that they are drawn from the binomial distribution. Assuming the beta prior and the binomial likelihood, it is possible to estimate the closed-form Bernoulli posterior distributions.

(3) Using the *bayesAB* R package[11], we estimated the posterior distributions for each of the two popu-

---

[11]https://CRAN.R-project.org/package=bayesAB

lations (younger and older) and calculated the probability that the sentiment scores of younger MPs are higher than the scores of older MPs.

The results show certain differences between the countries, In Bulgaria, Spain, and France, the MPs between 50 and 65 express the most positive sentiment, i.e. the thresholds for 55, 60, and 65 show high probabilities that younger MPs express positive sentiment while for the threshold of 50, this probability is lower. In the Czech Republic the positive sentiment is prominent in the age group of less than 55. In Slovenia, the MPs between 50 and 55 are the most negative, while other age groups are predominantly positive.

Table 14

The probabilities that MPs younger than the Age Cutoff express more positive sentiments compared to older ones, as calculated with the Bayesian AB test. For UK parliament the age of MPs is not available.

Age Cutoff	BG	CZ	ES	FR	SI	UK
50	61.2	83.8	27.3	16.5	88.1	/
55	91.1	98.8	73.6	73.6	27.0	/
60	95.5	54.9	100	95.6	94.1	/
65	99.9	4.9	97.1	87.0	99.9	/
70	73.7	0.8	7.8	73.3	94.1	/
75	64.9	7.1	2.6	15.6	74.1	/

5.4. Differences in political wing distributions

In this section, we further discuss some interesting questions related to the language of extreme left and right-wing politicians, i.e. we try to discover the language differences in political orientation in relation to age and gender. As explained in Section 4.1, we separated political parties according to their political orientation (extreme left or extreme right), based on the publicly available information and parties' self-declarations.

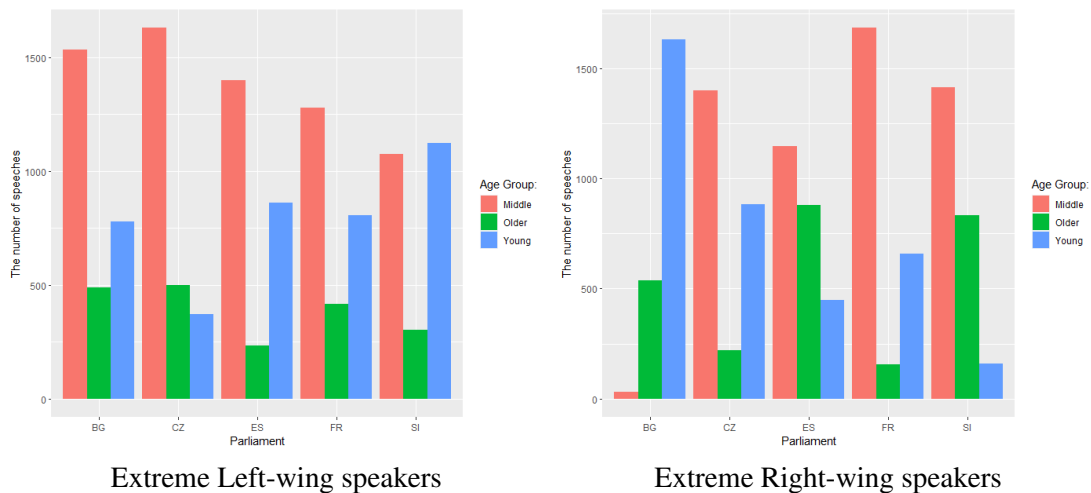Extreme Left-wing speakers   Extreme Right-wing speakers

Fig. 5. The distribution of the number of speeches relative to the MP's age group across the parliaments for extreme left- and right-wing MPs.

We first plot the differences in political orientation based on the age of speakers. Similarly to Section 5.1.1, we split the age span into three intervals, using the first and third quartile as the thresholds: younger MPs (less than the first quartile), middle-aged MPs (between the first and third quartile), and older MPs (more than the third quartile). The distribution plots for left- and right-wing speeches are presented in Figure 5. We can observe that for most of the parliaments, the extreme-left speakers are predominantly middle-aged. The exception is Slovenia, where both younger and middle-aged speakers form the the majority of extreme left-wing speakers. The extreme-right politicians in most parliaments are also predominantly middle-aged, except in Bulgaria, where the younger MPs for the majority of this group.

Similarly, we compared gender differences for both extreme-left and extreme-right positioned speakers in Figure 6. The difference in gender distributions are the most pronounced in Bulgaria and France. While for Bulgaria the number of female speakers drops going from extreme left to extreme-right, for France this number increases.
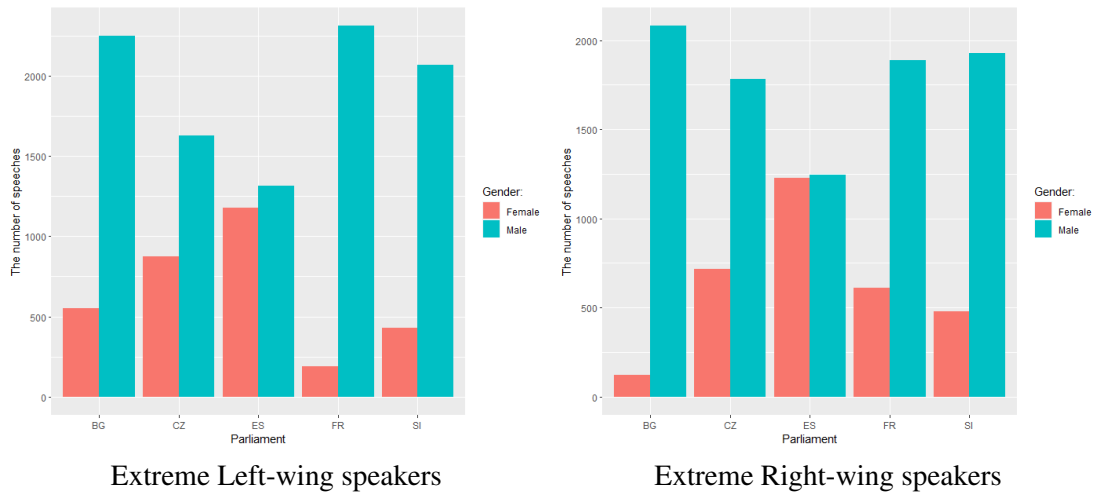
Extreme Left-wing speakers   Extreme Right-wing speakers

Fig. 6. The distribution of the number of speeches relative to the MP's age group across the parliaments for left- and right-wing MPs.

6. Conclusions and Further Work

We presented the mono- and cross-lingual methodology based on the cross-domain transfer learning and Bayesian statistical testing for the analysis of parliamentary speeches. The proposed methodology and constructed models can be applied in a uniform way to the parliamentary speeches in the ParliaMint corpora collection (and other parliamentary datasets with similar information). Our methodology covers analysis of sentiment and emotions, as well as prediction of metadata such as age, gender, and political orientation of the speakers. The source code of the developed methods and evaluation scenarios is publicly available[12]. We demonstrate the presented methodology on six national parliaments showing similarities and some surprising differences between them.

---

[12]https://github.com/KristianMiok/Parliamentary-Discourse

We discovered that in all countries except France, the age and gender of speakers is a strong factor in the political discourse. Further, we found a big difference in the discourse between extreme left- and right-wing parties in all analyzed countries. Surprisingly, there is also a considerable difference between center-left and center-right parties in all countries except France. The sentiment analysis shows considerable differences between parliaments. The Czech, Spanish and United Kingdom parliaments express less negative than positive sentiment, the Bulgarian and French parliaments have a balanced distribution, and in the Slovenian parliament, the negative sentiment dominates. The sentiment is also significantly different for different age groups. The situation is different with the emotions, where positive emotions are strongly dominant in all countries except France and UK, where positive and negative emotions are almost balanced.

There are many open avenues for further work. A larger analysis of all 16 parliaments in the ParliaMint collection would require a much larger research team who would be able to interpret the results but would produce a very interesting comparison between the parliaments. The proposed methodology could be extended with better training datasets for sentiment and emotions when they become available. We could analyze a broader spectrum of emotions, but currently, existing datasets are inadequate for our purpose due to differences in the covered domains.

Acknowledgements

References

[1] Abercrombie, G. and Batista-Navarro, R. T. (2020). ParlVote: A corpus for sentiment analysis of political debates. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5073–5078.

[2] Alba-Juez, L. (2018). Emotion and appraisal processes in language. *The Construction of Discourse as Verbal Interaction, Amsterdam, John Benjamins*, pages 227–250.

[3] Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 579–586.

[4] Argamon, S., Koppel, M., Fine, J., and Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text & Talk*, 23(3):321–346.

[5] Baraniak, K. and Sydow, M. (2021). A dataset for sentiment analysis of entities in news headlines (SEN). *Procedia Computer Science*, 192:3627–3636.

[6] Barrett, L. F. (2020). *Seven and a Half Lessons about the Brain*. Houghton Mifflin Harcourt Boston.

[7] Betz, H.-G. (1993). The new politics of resentment: Radical right-wing populist parties in Western Europe. *Comparative politics*, pages 413–427.

[8] Bučar, J., Žnidaršič, M., and Povh, J. (2018). Annotated news corpora and a lexicon for sentiment analysis in Slovene. *Language Resources and Evaluation*, 52(3):895–919.

[9] Cameron, R. (2011). Aging, age, and sociolinguistics. *The handbook of Hispanic sociolinguistics*, page 207.

[10] Coates, J. (2015). *Women, men and language: A sociolinguistic account of gender differences in language*. Routledge.

[11] Cortal, G., Finkel, A., Paroubek, P., and Ye, L. (2022). Natural language processing for cognitive analysis of emotions. *arXiv preprint arXiv:2210.05296*.

[12] Dahllöf, M. (2012). Automatic prediction of gender, political affiliation, and age in Swedish politicians from the wording of their speeches—A comparative study of classifiability. *Literary and linguistic computing*, 27(2):139–153.

[13] de Féral, C. (2012). "youth languages": A useful invention? *Langage et societe*, 141(3):21–46.

[14] De la Torre, C. (1997). Populism and democracy: political discourses and cultures in contemporary Ecuador. *Latin American Perspectives*, 24(3):12–24.

[15] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Nemade, G., and Ravi, S. (2020). Goemotions: A dataset of fine-grained emotions. *arXiv preprint arXiv:2005.00547*.

[16] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

[17] Dzieciatko, M. (2018). Application of text analytics to analyze emotions in the speeches. In *International Conference on Information Technologies in Biomedicine*, pages 525–536. Springer.

[18] Elkink, J. A. and Farrell, D. M. (2021). Predicting vote choice in the 2020 Irish general election. *Irish Political Studies*, 36(4):521–534.

[19] Emara, I. (2017). Gender identity construction in Facebook statuses of Egyptian young adults. *Cairo Studies in English*, 2017(1):86–111.

[20] Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubesic, N., Simov, K., Pancur, A., Rudolf, M., Kopp, M., Barkarson, S., Steingrmsson, S., et al. (2022). The ParlaMint corpora of parliamentary proceedings. *Language resources and evaluation*, pages 1–34.

[21] Flaherty, P. (1987). Langue nationale/langue naturelle: The politics of linguistic uniformity during the French revolution. *Historical Reflections/Réflexions historiques*, pages 311–328.

[22] Frizelle, P., Thompson, P. A., McDonald, D., and Bishop, D. V. (2018). Growth in syntactic complexity between four years and adulthood: evidence from a narrative task. *Journal of Child Language*, 45(5):1174–1197.

[23] Ghafournia, N. (2015). Language as a symbol of group membership. *Asian Social Science*, 11(5):19.

[24] Glavaš, G., Nanni, F., and Ponzetto, S. P. (2017). Unsupervised cross-lingual scaling of political texts. In *EACL 2017*, page 688.

[25] Han, J.-H. (2007). Analysing roll calls of the European Parliament: A Bayesian application. *European Union Politics*, 8(4):479–507.

[26] Hansen, M. E. (2009). The positions of Irish parliamentary parties 1937–2006. *Irish Political Studies*, 24(1):29–44.

[27] Hidalgo-Tenorio, E. (2016). Genderlect. *The Wiley Blackwell encyclopedia of gender and sexuality studies*.

[28] Honkela, T., Korhonen, J., Lagus, K., and Saarinen, E. (2014). Five-dimensional sentiment analysis of corpora, documents and words. In *Advances in self-organizing maps and learning vector quantization*, pages 209–218. Springer.

[29] Hopkins, D. J. and King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 54(1):229–247.

[30] Iliev, I. R., Huang, X., and Gel, Y. R. (2019). Political rhetoric through the lens of non-parametric statistics: are our legislators that different? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(2):583–604.

[31] Jay, K. L. and Jay, T. B. (2013). A child's garden of curses: A gender, historical, and age-related evaluation of the taboo lexicon. *The American Journal of Psychology*, 126(4):459–475.

[32] Kerswill, P. (1996). Children, adolescents, and language change. *Language variation and change*, 8(2):177–202.

[33] Kowsari, K., Heidarysafa, M., Odukoya, T., Potter, P., Barnes, L. E., and Brown, D. E. (2020). Gender detection on social networks using ensemble deep learning. In *Proceedings of the Future Technologies Conference*, pages 346–358.

[34] Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania press.

[35] Lakoff, R. (1973). Language and woman's place. *Language in society*, 2(1):45–79.

[36] Lehti, L. and Laippala, V. (2014). Style in French politicians' blogs: Degree of formality. *Language internet*, 11(1).

[37] Malo, P., Sinha, A., Korhonen, P., Wallenius, J., and Takala, P. (2014). Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology*, 65(4):782–796.

[38] Maurer, A. and Wessels, W. (2001). *National parliaments on their ways to Europe. Losers or latecomers?* Nomos Verlag.

[39] Menéndez, D. A., González-Barahona, J. M., and Robles, G. (2020). Damegender: Writing and comparing gender detection tools. In *Proceedings of the Seminar Series on Advanced Techniques & Tools for Software Evolution, SATToSE*.

[40] Milroy, J. and Milroy, L. (1985). Linguistic change, social network and speaker innovation1. *Journal of linguistics*, 21(2):339–384.

[41] Miok, K., Nguyen-Doan, D., Škrlj, B., Zaharie, D., and Robnik-Šikonja, M. (2019). Prediction uncertainty estimation for hate speech classification. In *Statistical Language and Speech Processing: 7th International Conference, SLSP 2019, Ljubljana, Slovenia, October 14–16, 2019, Proceedings 7*, pages 286–298. Springer.

[42] Miok, K., Škrlj, B., Zaharie, D., and Robnik-Šikonja, M. (2021). To BAN or not to BAN: Bayesian attention networks for reliable hate speech detection. *Cognitive Computation*, pages 1–19.

[43] Montalvo, J. G., Papaspiliopoulos, O., and Stumpf-Fétizon, T. (2019). Bayesian forecasting of electoral outcomes with new parties' competition. *European Journal of Political Economy*, 59:52–70.

[44] Murphy, B. (2010). *Corpus and sociolinguistics: Investigating age and gender in female talk*, volume 38. John Benjamins Publishing.

[45] Oberländer, L. A. M. and Klinger, R. (2018). An analysis of annotated corpora for emotion classification in text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119.

[46] Öhman, E., Pàmies, M., Kajava, K., and Tiedemann, J. (2020). XED: A multilingual dataset for sentiment analysis and emotion detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6542–6552.

[47] Park, S. and Woo, J. (2019). Gender classification using sentiment analysis and deep learning in a health web forum. *Applied Sciences*, 9(6):1249.

[48] Rheault, L., Beelen, K., Cochrane, C., and Hirst, G. (2016). Measuring emotion in parliamentary debates with automated textual analysis. *PloS ONE*, 11(12):e0168843.

[49] Robnik-Šikonja, M., Reba, K., and Mozetič, I. (2021). Cross-lingual transfer of sentiment classifiers. *Slovenščina 2.0: empirical, applied and interdisciplinary research*, 9(1):1–25.

[50] Rodriguez Gonzalez, F. (1994). Youth and student slang in british and american english: an annotated bibliography. *Revista alicantina de estudios ingleses*, 7:201–212.

[51] Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., and Sedlmair, M. (2017). Supervised sentiment analysis of parliamentary speeches and news reports. In *67th Annual Conference of the International Communication Association (ICA), Panel on Automatic Sentiment Analysis*.

[52] Sailunaz, K., Dhaliwal, M., Rokne, J., and Alhajj, R. (2018). Emotion detection from text and speech: a survey. *Social Network Analysis and Mining*, 8:1–26.

[53] Sakamoto, T. and Takikawa, H. (2017). Cross-national measurement of polarization in political discourse: Analyzing floor debate in the us the japanese legislatures. In *2017 IEEE international conference on big data*, pages 3104–3110.

[54] Salmela, M. and Von Scheve, C. (2017). Emotional roots of right-wing political populism. *Social Science Information*, 56(4):567–595.

[55] Saravia, E., Liu, H.-C. T., Huang, Y.-H., Wu, J., and Chen, Y.-S. (2018). Carer: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3687–3697.

[56] Stenström, A.-B. and Jørgensen, A. M. (2009). *Youngspeak in a multilingual perspective*, volume 184. John Benjamins Publishing.

[57] Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban british english of norwich. *Language in society*, 1(2):179–195.

[58] Van Dijk, T. A. (1985). Handbook of discourse analysis. In *Discourse and dialogue*.

[59] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.