UNIVERSITY OF COPENHAGEN
DEPARTMENT OF NORDIC STUDIES AND LINGUISTICS

Master Thesis
IT & Cognition

Domain-Specific Sentiment Analysis

A Comparative Study of Two Parliaments

Yana Nikolova
xjv550@alumni.ku.dk

May 31, 2023
Supervisor: Costanza Navarretta
Characters: 133 250 (55,5 pages)

ABSTRACT

In recent years, the availability of parliamentary transcriptions has motivated the use of Natural Language Processing (NLP) methods for investigating these datasets quantitatively. One area of study for political scientists has been the expression of emotion by parliamentary speakers and the factors that influence it. However, there is a lack of research on non-anglophone legislatures and insufficient validation of existing sentiment analysis methods. In this study, a technique for creating domain-specific sentiment lexica with word embeddings is applied to datasets of parliamentary speech from Bulgaria and Denmark in order to investigate the effects of party status and proceeding type. By embedding each corpus with Global Vectors (GloVe), manually produced seed lists of positive and negative words can be expanded using measures of vector similarity. The resulting lexica, called ParLex DK/BG, are used to score and compare the two corpora. The sentiment analysis indicates that for both legislatures, government/coalition status is associated with more positive emotion relative to opposition or other status, though this is much more pronounced in Denmark. With regard to proceeding type, there is no consistent cross-national pattern, but proceeding type does interact with party status. In both countries, sentiment gaps between coalition and opposition are largest during ministerial questionings, and in the Bulgarian parliament, the effect of party status is only clearly apparent during questionings and draft decisions. These differences may reflect the contrast between a more established (Denmark) and more politically turbulent (Bulgaria) legislature. Confidence in these results is also moderated by evaluation with a hand-annotated gold standard of speeches. Correlation between ParLex and manual annotation is 0.40 for Bulgarian and 0.48 for Danish. ParLex DK is also outperformed by a common general-use lexicon, indicating that domain-specificity does not offer clear advantages when analyzing parliamentary speech. Both lexica are made publicly available with suggestions and recommendations for future work.

Keywords: NLP, sentiment analysis, sentiment lexicon, GloVe, parliament, opinion mining, Bulgaria, Denmark

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Motivation

As part of the decision-making apparatus of modern democratic states, parliamentary speeches and their political and legislative outcomes impact the lives of millions of people. Members of parliament discuss and vote on bills, budgets and taxes, debate issues of national interest and exercise oversight over the executive authority – largely through language. Understanding the dynamics of parliamentary speech can therefore provide crucial insight into the political engine that runs a country. Parliamentary transcriptions are often publicly accessible through government websites, although formatting and searchability issues, as well as the sheer volume of textual output, have pushed researchers studying this data toward computational solutions with Natural Language Processing (NLP). Techniques for automated textual analysis have been used to investigate everything from gendered speaking patterns (Hargrave & Blumenau, 2022) to political polarization (Fiva et al., 2021) in parliaments.

One strand of this work has focused on the emotional and affective character of parliamentary speech, integrating insights and methods from the field of sentiment analysis (Abercrombie & Batista-Navarro, 2020). In parliament, emotions influence behavior, shape rhetorical strategies and express attitudes. Political roles, such as government and opposition, can be investigated and indexed in terms of emotional polarity (Abercrombie & Batista-Navarro, 2018; Proksch et al., 2019). Furthermore, different proceeding types and parliamentary contexts lend themselves to divergent expressions of emotion (e.g., Rudkowsky et al., 2018).

1.2 Overview

In this thesis, such questions will be explored in a comparative framework, contrasting the parliaments of Bulgaria and Denmark. While both countries are parliamentary democracies, they differ widely in both political culture and socio-economic conditions, presenting a suitable test case for the ability of sentiment analysis methods to characterize different parliamentary dynamics. This potential, as well as the author's personal relationship to the countries and knowledge of their languages, is why they were chosen for comparison.

Moreover, the literature reveals a relative paucity of research into non-anglophone legislatures, especially with regard to sentiment, making this work an important extension of existing knowledge in political science. The approach chosen for estimating emotional polarity in the two parliaments – the creation and implementation of domain-specific sentiment dictionaries using word embeddings – will also serve as a contribution to the under-resourced Danish and Bulgarian NLP sub-fields and a partial replication of the methodology in Rheault et al. (2016). The resulting sentiment lexica, dubbed ParLex BG/DK, have been made available on GitHub[1] for use by other interested parties.

Two main factors will be considered in their influence on expressed sentiment: party status and proceeding type. The effect of party status is quite well-established in the literature, which shows that opposition speakers tend to express more negative emotion than government/coalition speakers (Proksch et al., 2019; Rheault et al., 2016; Rudkowsky et al., 2018). The replicability and magnitude of this effect across both parliaments will therefore be of interest.

On the other hand, the effect of different proceeding types in parliament – bill readings, ministerial hearings, negotiations – is understudied, with an assumption (e.g., in parts of Proksch et al., 2019) that the most parliamentary speech is dedicated to legislative debate. The relative prominence of different proceeding types and their effect on sentiment patterns among speakers of different party status are therefore central questions in this comparative framework.

Finally, a main concern of this study is the usefulness of such analysis. While the development and adoption of computational techniques in fields such as political science has been enthusiastic, concerns about the validity and analytical scope of these methods abound (Grimmer & Stewart, 2013).

[1] https://github.com/yananikol/parlex

Performing a rigorous evaluation of the methodology and assessing the interpretability of the results will therefore make a broader contribution to the question of the applicability of NLP approaches in the humanities.

1.3 Research Problem

To summarize, the objective of this thesis is to answer the following questions:

How can emotional polarity in the Bulgarian and Danish parliaments be modelled and compared cross-linguistically? And what comparative insights can it provide into the political and legislative dynamics of parliamentary discourse in the two countries?

This problem can be subdivided into four research questions:

I. How can sentiment analysis be adapted to parliamentary speech in order to measure emotional polarity?
II. How does the use of emotionally charged language compare across party status (government, opposition, neither)?
III. How does the use of emotionally charged language compare across different types of parliamentary speech (debates, question time, etc.)?
IV. Do the same patterns hold in both parliaments? What do differences imply about the political and legislative cultures in Denmark and Bulgaria?

In addressing these questions, the thesis will explore and extend state-of-the-art NLP techniques for sentiment analysis to gain insight into the linguistic and political dynamics of two very different European countries.

1.3.1 Hypotheses

While this study has an exploratory dimension, previous work and prior knowledge allow for the formulation of hypotheses relating to each research question. Fuller justification for these hypotheses is provided in Chapters 2 (for Hypothesis II., III. and IV.) and 4 (for Hypothesis I.).

I. Word embeddings can be used to extend initial seed lists of sentiment words with domain-specific terms for a more accurate sentiment analysis than existing general-use sentiment lexica can provide.

II. Consistent with their role as critics of the cabinet, opposition speakers are expected to exhibit less positivity than coalition/government speakers.
III. Because of their adversarial character, speeches made during hearings and questioning of ministers are expected to score less positively on average than speeches made during legislative debates.
IV. The turbulent character of Bulgarian parliamentary politics relative to the Danish politics is expected to result in a) more pronounced differences between coalition and opposition, b) more negative sentiment, and c) more emotional language in general.

The hypotheses are evaluated in Chapter 6 with a summary in Chapter 7.

1.4 Thesis Structure

The structure of the thesis is as follows: Chapter 2 reviews existing literature on automated textual processing of parliamentary speech with a focus on sentiment analysis. Relevant background information and prior linguistic and sociological work on the Danish and Bulgarian parliaments is also presented. Chapter 3 introduces the ParlaMint datasets and the process of data formatting and preparation, including additional feature extraction and the creation of a gold standard for evaluation. This chapter also provides a descriptive statistical overview of both parliamentary datasets.

The inspiration and theory behind the methodology for domain-specific sentiment analysis are explained in Chapter 4, as well as the practical details and challenges of implementation for both languages. In Chapter 5, the results of the sentiment analysis are explored through a qualitative sentiment keyword analysis and a quantitative comparison of sentiment scores across party status, gender, and proceeding type for both parliaments. These results are evaluated in Chapter 6 by assessing the construct validity of sentiment in relation to the gold standard. This chapter further interprets and discusses findings in light of previous research and hypotheses.

Finally, Chapter 7 summarizes the substantive and methodological contributions of the thesis, issuing recommendations and suggestions for future work in the field.

2 BACKGROUND

This chapter motivates the hypotheses and overall approach of the thesis in light of previous work. Section 2.1 presents related work on the use of NLP techniques for analyzing political speech and measuring political difference. Section 2.2 is an introduction to the two parliaments, covering basic structure, demographic composition and existing research on their language and discourse.

2.1 Related Work

Because of their availability and political relevance, parliamentary speeches have been studied within a wide variety of fields and theoretical perspectives – linguistics, political science, sociology, and more (e.g., Diermeier et al., 2012; Gregersen, 2019; Ilie, 2015; Osenova, 2022). Traditionally, sociological work on parliamentary speech has focused on critical discourse and content analysis with manual coding and qualitative analysis (Skubic & Fišer, 2022), but more recent quantitative approaches – initially termed "text-as-data" among political scientists (Grimmer & Stewart, 2013) – have deployed NLP techniques to yield new methodologies and insights.

One of the main strands of this work, reviewed in detail in Abercrombie & Batista-Navarro (2020), has used parliamentary speech data for sentiment analysis write large – emotion analysis, position scaling, party detection, opinion-topic analysis, etc. – in short, the quantification of subjective states based on linguistic material. In the following sections, some major strokes of this work will be reviewed and contextualized in relation to the current study.

2.1.1 Political Difference in Parliament

Much of the work on opinion mining for parliamentary texts has revolved around the question of political difference. Researchers have sought to extract, quantify, and classify political positions and identities from text, with early scaling methodologies such as Wordscores (Laver et al., 2003) or Wordfish (Slapin & Proksch, 2008) giving way to machine learning (Goet, 2019) and, recently, neural network (Contreras et al., 2022; Han, 2022) approaches. Unsupervised techniques like Latent Dirichlet Allocation (Blei et al., 2003) are popular for topic analysis in particular (Chen et al., 2023; Contreras et al., 2022; Müller-Hansen et al., 2021), but most other tasks involve some kind of supervised machine classification. Support Vector Machines and Naïve Bayes Classifiers are the most popular classifier choices (Abercrombie & Batista-Navarro, 2020) and have been used to classify parliamentary speakers by party affiliation (Dahllöf, 2012; Hansen et al., 2019; Hirst et al., 2010; Navarretta & Hansen, 2020; Yu et al., 2008), age (Dahllöf, 2012) and gender (Dahllöf, 2012; Mandravickaitė & Oakes, 2016), helping elucidate the socio-linguistic basis of such differences. Others have used classification accuracy as a quantity of interest to measure political polarization (Peterson & Spirling, 2018; Søyland & Lapponi, 2017), drawing on Bayesian insights about political dissimilarity (Gentzkow et al., 2016).

However, a challenge to classification and scaling approaches to parliamentary speech has been the lack of attention paid to the institutional structure of debates. Speeches given by different speakers and groupings are not necessarily commensurable, as taking the floor in parliament is decided by factors as diverse as gender expectations (Koppensteiner et al., 2016) and party committees (Proksch & Slapin, 2012). Consequently, MPs are not necessarily setting their own agendas, and their language is not purely a function of their own political positions and priorities. For example, a party classification study by Hirst et al. (2010) points out that language of "attack and defense" – contingent on party status as coalition or opposition – explains a large chunk of the variation in party language, acting as a confounder.

A number of studies examining emotion and sentiment polarity in parliaments have emphasized this point. A major finding in this area is the consistent difference in average sentiment polarity between government and opposition parties (Proksch et al., 2019; Rheault et al., 2016; Rudkowsky et al., 2018). Another issue is that government-led committees or cabinets propose the bulk of the legislation in most parliaments, setting the agenda and topics of discussion. This limits the usefulness of topics or topic-related vocabulary (Bhatia & P, 2018; Sakamoto & Takikawa, 2017) for party classification, especially without the incorporation sentiment or stance. Indeed, Søyland & Lapponi (2017) report that, based on classification accuracy metrics, the most and least immigration-friendly Norwegian parties appear most similar to each other on that topic – simply because they speak about it often and therefore share a substantial vocabulary. Accordingly, sentiment is crucial to understanding the dynamics of parliamentary speech and a pre-condition to performing more complex and wide-ranging analyses of political difference.

2.1.2 Measuring Emotion in Parliament

Studies estimating emotion in parliamentary speech most commonly rely on simple word-counting methods using dictionaries (or lexica) of labelled sentiment words, often incorporating syntactic rules to account for negation and intensity (Onyimadu et al., 2014; Rheault et al., 2016). These dictionaries can be extended with domain-specific terms, for example by using word embeddings (Hargrave & Blumenau, 2022; Rheault et al., 2016; Vries, 2022). Rudkowsky et al. (2018), on the other hand, use word embeddings to compare sentences from the Austrian parliament to a hand-annotated set of embedded political speeches and thereby determine sentiment. Others, like Abercrombie (2021) or Mochtak et al. (2022) create large sentiment-annotated corpora of parliamentary speeches and train machine learning algorithms for classification.

Once generated, sentiment scores are used and analyzed in different ways. Abercrombie (2021) and Bhatia & P (2018) combine sentiment and topic to automatically detect political stances, while Bor et al. (2023) use topic modelling to create topical sub-corpora and apply sentiment analysis to identify areas of political polarization. Proksch et al. (2019) use sentiment divergence between coalition and opposition speakers as a measure of legislative conflict and a predictor of bill voting outcomes. Others (Hargrave & Blumenau, 2022; Rheault et al., 2016; Rudkowsky et al., 2018) test hypotheses about the effect of speaker roles, proceeding types, gender, or economic forces on sentiment in parliament to better understand the socio-linguistic forces shaping parliamentary discourse. For example, Rudkowsky et al. (2018) find that urgent question debates in the Austrian parliament are associated with greater negativity than other proceeding types – forming the background for Hypothesis III.

This thesis explores sentiment in a similar vein, investigating the effects of party status and proceeding type. Following Rheault et al. (2016) closely in methodology, the goal is to extend and evaluate existing NLP research on sentiment analysis by applying it to novel and under-resourced languages and contexts – the Bulgarian and Danish parliaments.

2.2 National Parliaments

2.2.1 Denmark

The Danish Parliament or People's Assembly (Danish: *Folketing*) is the national legislature of the Kingdom of Denmark, comprising Denmark and the semi-autonomous territories Greenland and the Faroe Islands. Its 179 MPs are elected by proportional representation. A chairperson (Danish: *formand*) is responsible for order and supervision during parliamentary sessions.

Character and Composition

Government and coalition-building in the Danish parliament is highly cooperative, with minority governments being the rule rather than the exception since WWII. In this system of "negative parliamentarism", a government need only be tolerated, rather than directly supported, by a majority in parliament, which incentivizes collaboration across party and government lines (Damgaard, 2004).

Like other established proportional democracies like Germany, political parties in Denmark display strong party discipline in voting patterns and ideological positioning (Green-Pedersen & Skjæveland, 2020), relying strongly on "party branding" during elections as well (Elmelund-Præstekær & Schumacher, 2014). While the average seniority of MPs has been falling in recent years (Christiansen, 2017), there is still a strong trend of reelecting incumbents, reinforcing institutional and ideological stability within parties.

Bloc discipline is also high, as parties can be reliably grouped into voting blocs along a left-right spectrum for both traditional political issues such as redistribution and social reform and newer political issues such as EU integration, refugee policy and globalization (Green-Pedersen & Skjæveland, 2020).

In terms of gender composition, the female share of parliament has meandered between 35-40% with a slight upward trend since 1990, surpassing 40% for the first time in 2022 (Folketinget, 2023).

Language and Discourse

Meetings and debates in the Danish Parliament proceed in a formal tone, including the use of otherwise obsolete honorifics and a mandatory third person address when referring to other MPs (Folketinget, 2017). Because the parliamentary agenda is set in advance, ministers, and party foremen (Danish: *ordførere*) will read from prepared manuscripts or detailed notes during bill readings, while supplementary remarks and debates during readings and hearings are unprepared and spontaneous, in closer alignment with a colloquial register (Gregersen, 2019).

Most work with Danish parliamentary speech data has been in the realm of traditional discourse analysis, interrogating discursive strategies and how they structure political and social conflicts in parliament – e.g., debates on immigration and refugees (Holm, 2006) or artificial insemination (Petersen, 2009).

In recent years however, the proliferation of computational methods and resources for analyzing text have inspired a number of studies using this material, particularly among researchers associated with the ParlaMint project (Erjavec et al., 2022) and ParlaCLARIN conferences. For example, Hansen et al. (2019) use lexical and metadata features for automated topic classification of parliamentary speeches, while. Hansen et al. (2018) investigate speech patterns and frequency by gender. Finally, Navarretta & Hansen (2020) use frequency analyses to characterize four party manifestos and attempt to predict party membership from parliamentary speeches using lexical features and a variety of machine learning classifiers. However, no studies on sentiment or emotion in the Danish parliament could be found, indicating that this work fills an important research gap.

2.2.2 Bulgaria

The National Assembly (Bulgarian: *Народно събрание*, romanized: *Narodno sabranie*) is the main legislative body of Bulgaria, comprising 240 MPs who are elected by proportional representation to serve 4-year terms. Parliamentary sessions are presided over by a chairperson (Bul.: *председател*).

Character and Composition

The National Assembly is characterized by a high degree of party fluidity and fragmentation, similarly to other post-communist legislatures in Central and

Eastern Europe (Kolarova, 2018). Party and coalition fragmentation, as well as frequent corruption scandals, often lead to votes of no confidence and premature elections, cutting legislative periods short. This is related to the fact that Bulgarian parties tend to be organized around charismatic leaders and populist messaging rather than well-developed ideologies (Karasimeonov, 2019). Personal conflicts and sharp confrontations abound, as evidenced in both antagonistic discourse (Tarasheva, 2014) and party fragmentation. Between 1991 and 2017, an average of 5,3% of sitting MPs left their parties to become independents (Kolarova, 2018).

Other unusual features of parliamentary composition in Bulgaria include the consistently high turnover of MPs across legislative periods (Kolarova, 2018): 40-70% of MPs in the 8 legislative periods between 1991-2017 were serving their first ever term. This contributes to party and ideological instability and might result in less stable discursive and rhetorical norms.

Female representation in the National Assembly has converged to 20-25% in recent years. Newer parties tend to have higher proportions of female MPs according to Kolarova (2018).

Language and Discourse

Parliamentary speech data from the National Assembly has been employed for linguistic (Osenova, 2022a, 2022b) and discourse analytical (Mavrodieva, 2014) studies, but it has also been analyzed using methods from NLP. Tarasheva (2004) performs a comparative keyword analysis of the January 2009 plenary sessions of the Bulgarian and UK parliaments, finding both topical and procedural differences between the two. The Bulgarian parliament is characterized by shorter sessions, more procedural language, more internal (vs. foreign) topics of discussion and a greater degree of formality than the UK parliament.

Another keyword and collocation study by Tarasheva (2014) examines a corpus of transcriptions from the turbulent and short-lived 42nd National Assembly (2013-2014). The parliamentary speeches show ample evidence of antagonism: 2nd person pronouns appear often in the context of confrontations and accusations. Party and personal names, as well as direct addresses, also feature extensively, highlighting the internal nature of the conflicts. Procedural language is conspicuous, as well as references to noise and exclamations/shouts (Bulgarian: *възгласи*).

Finally, Miok et al. (2022) analyze the Bulgarian ParlaMint corpus (see Section 3.1) in a wide-ranging comparative study. They find that public financing, infrastructure, health, corruption and procedural debates between coalition and opposition parties are prominent topics. A sentiment analysis on randomly sampled speeches indicates that the proportion of positive and negative speech is roughly equal.

2.2.3 Summary of Contrasts

Tab. 2.1 summarizes the main differences between the two parliaments. The adversarial and turbulent character of the Bulgarian parliament is hypothesized to be associated with more overall negative sentiment and large sentiment differences between coalition and opposition parties (Hypothesis IV). Procedural language may result in a larger proportion of neutral speeches as well. On the other hand, the cooperative and disciplined Danish style of parliamentary debate is expected to result in smaller, but potentially more consistent differences in sentiment between coalition and opposition.

Feature	Denmark	Bulgaria
Party Discipline	High	Low
Procedural Language	-	High
Female Representation	30-40%	20-25%
MP Background	Predominantly incumbent	Predominantly novice
Style	Cooperative	Adversarial
Government	Minority	Majority

Table 2.1 Summary of differences between the Danish and Bulgarian Parliament

3 DATA

This chapter presents an overview of the data forming the empirical basis of the thesis. Section 3.1 explains the background and formatting of the ParlaMint corpora and motivates their use. Data preparation, including format conversions and feature extraction, is presented in Section 3.2. The creation and characteristics of an evaluative gold standard are laid out in 3.3. Finally, Section 3.4 provides a quantitative overview of both corpora across relevant metrics.

3.1 The ParlaMint Corpora

Many countries make transcriptions of their parliamentary proceedings publicly available, but there is little uniformity in formatting and available metadata information, making comparison and common use difficult. This is the problem that the ParlaMint project, associated with the European research infrastructure CLARIN (Common Language Resources and Technology Infrastructure), aims to address. The project has engaged researchers from all over Europe to collect and annotate the parliamentary speech data according to a common standard, the Parla-CLARIN TEI-based encoding (Erjavec et al., 2022). The first stage of ParlaMint (July 2020 – May 2021) produced 17 national corpora, among them a version of the Danish and Bulgarian corpora used in this project. ParlaMint II, which is ongoing at the time of writing, will extend existing corpora with new data and add create corpora for missing European legislatures, while updating and enhancing annotations. This extended data (version 3.0) contains proceedings up to 2022 and was provided to the author by the responsible ParlaMint researchers via Costanza Navarretta, the supervisor of this thesis, as it is not yet published on the CLARIN online repository (Erjavec et al., 2021).

3.1.1 Objective

Using the ParlaMint data will bypass the time-consuming process of scraping, cleaning, and processing parliamentary proceedings and metadata from government websites, permitting a more extensive and fine-grained analysis. The standardized TEI format of ParlaMint also greatly facilitates the pre-processing that is still necessary, as most custom-written code can be reused or easily adapted to the different national corpora, expanding the range of analysis and comparison.

One drawback of the ParlaMint datasets is the relatively short time period they cover – around 8 years between 2014-2022, corresponding to only 2-3 legislative periods for most countries. Previous work on parliamentary sentiment like Rheault et al. (2016) encompasses larger timescales and investigates long-term trends. Instead, the goal with ParlaMint is explore more deeply the different contexts of parliamentary speech and evaluate results comprehensively through the creation of gold standard. The smaller timescale ensures that even a modest gold standard can be representative and contribute to a meaningful error analysis.

Consequently, the objectives of the data collection and preparation process are 1) the transformation of the corpora to a format suitable for analysis with a sentiment lexicon, containing all necessary metadata for answering the research questions and 2) the creation of a gold standard for evaluation.

3.1.2 Encoding and Annotation

The ParlaMint corpora are encoded and formatted as XML files, according to a schema based on TEI (Text Encoding Initiative), a widely used international and interdisciplinary encoding standard (TEI Consortium, 2023). A ParlaMint corpus exists in two TEI versions, one of which encodes speech segments as regular text, while the other provides detailed linguistic annotation of tokenized speech – noting lemma, part-of-speech, syntactic role, and morphological features for each token. Punctuation and named entity types are also encoded.

Each corpus comes with a root file, ParlaMint-XX.xml (XX being the country code), which contains the corpus header and a list of all component files. A number of other XML files provide taxonomies of the types of parliamentary meetings, lists of all parties and speakers in the corpus, ruling coalitions, etc. Party information includes active years, coalitions, and opposition status, while speaker encodings include name, sex, ministerial positions, and party affiliation. This information is used to generate metadata files with available conversion scripts (see 3.1.2).

Individual corpus TEI files generally encode a single continuous parliamentary meeting. A header provides general information about the ParlaMint corpus and researchers, as well as the term, session, and date of the meeting. The speeches themselves are identified by the speaker's name and role (regular, chair or guest), as well as a unique ID number. They are split into segments corresponding to transcription paragraphs. When linguistically annotated, these segments branch out into sentences and then individual tokens with annotation (Fig. 3.1). Each sentence is also parsed according to a dependency grammar.

```xml
...
<text xml:lang="da" ana="#reference">
      <body>
         <div type="debateSection">
            <head>Punkt 0</head>
            <note type="agendaItem">2014-10-09-0</note>
            <u who="#LykketoftMogens" xml:id="ParlaMint-DK_20141009100001" ana="#chair">
               <seg xml:id="ParlaMint-DK_20141009100001.seg1">
                  <s xml:id="ParlaMint-DK_20141009100001.seg1.1">
<w lemma="møde" msd="UPosTag=NOUN|Definite=Def|Gender=Neut|Number=Sing" xml:id="ParlaMint-
DK_20141009100001.seg1.1.1">Mødet</w>
<w lemma="være" msd="UPosTag=AUX|Mood=Ind|Tense=Pres|VerbForm=Fin|Voice=Act" xml:id="ParlaMint-
DK_20141009100001.seg1.1.2">er</w>
<w lemma="åbne" msd="UPosTag=VERB|Definite=Ind|Number=Sing|Tense=Past|VerbForm=Part" join="right"
xml:id="ParlaMint-DK_20141009100001.seg1.1.3">åbnet</w>
<pc msd="UPosTag=PUNCT" xml:id="ParlaMint-DK_20141009100001.seg1.1.4">.</pc>
<linkGrp targFunc="head argument" type="UD-SYN">
<link ana="ud-syn:nsubj" target="#ParlaMint-DK_20141009100001.seg1.1.3 #ParlaMint-
DK_20141009100001.seg1.1.1"/>
<link ana="ud-syn:aux" target="#ParlaMint-DK_20141009100001.seg1.1.3 #ParlaMint-
DK_20141009100001.seg1.1.2"/>
<link ana="ud-syn:root" target="#ParlaMint-DK_20141009100001.seg1.1 #ParlaMint-
DK_20141009100001.seg1.1.3"/>
<link ana="ud-syn:punct" target="#ParlaMint-DK_20141009100001.seg1.1.3 #ParlaMint-
DK_20141009100001.seg1.1.4"/>
</linkGrp>
                  </s>
...
```

Figure 3.1 Example of linguistically annotated sentence encoding from ParlaMint-DK.

3.1.3 Scripts and Transformation

This detailed markup provides valuable information but also renders the files difficult to use in their basic form. For this reason, the project researchers have developed and made available[2] a number of useful conversion and extraction scripts in XSLT and Perl. Functionalities include extracting metadata and speaker information, obtaining indexed plain-text speeches from TEI files, and converting annotated TEI files to concordance format. As the ParlaMint repository files (Erjavec et al., 2021), a newer version of which was provided to the author,

[2] https://github.com/clarin-eric/ParlaMint

actually include derived plain-text corpora with TSV metadata, it was not necessary to use the original text TEI corpora at any point.

With regard to the linguistically annotated TEI corpora, however, conversion scripts were of little use, as their formats are difficult to read for regular text processing purposes. Converting these to a usable format was a part of the data preparation process, summarized in Fig. 3.1 and described in the following section.

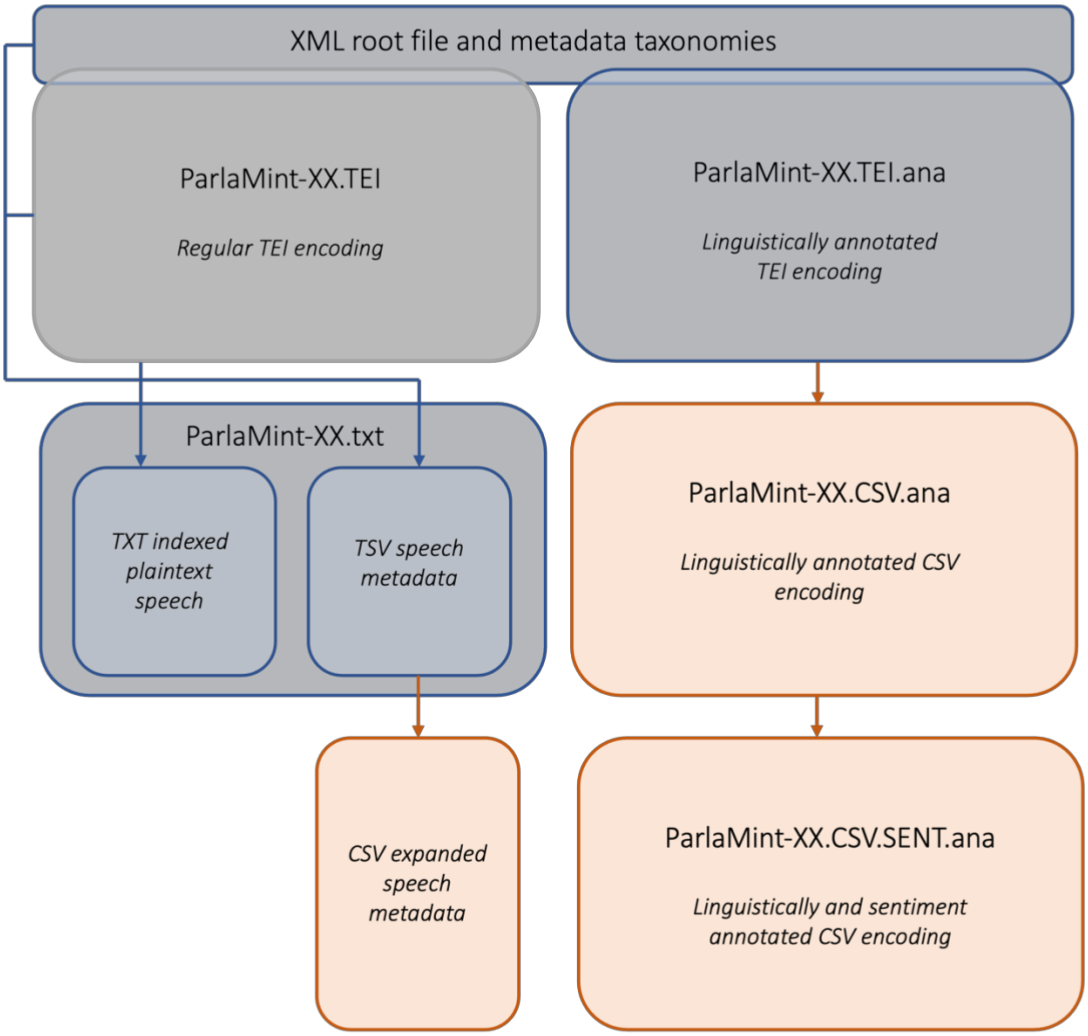

Figure 3.2 Schematization of data sources and pre-processing. Blue elements were obtained from the ParlaMint repository, while orange elements were produced through data preparation work. Elements in gray were not accessed for use. Arrows indicate processes of transformation or annotation.

3.2 Data Preparation

3.2.1 XML to CSV Conversion

To facilitate reading and processing with the common data analysis package Pandas (McKinney, 2010), the annotated TEI files were converted to a CSV format using the BeautifulSoup (Richardson, 2007) package for reading and traversing XML with Python. Although the ParlaMint TEI encoding is largely standardized, there were enough differences between the corpora (e.g., the format of speech ID's, placement of information about the number of words in a document, etc.) that it was necessary to write separate scripts for Danish and Bulgarian in order to extract the right information. The generated CSV files encode one word or punctuation mark per row, also noting the speech ID, the segment, sentence, and word numbers, named entity status, part-of-speech tag, and a variable number of other morphological markers (e.g., tense for verbs or gender for nouns) for each token. This CSV format was the basis for the sentiment analysis.

During sentiment tagging, the CSV files were expanded by one column to include a continuous sentiment score for each token, based on lookup in ParLex (see Section 4.2). This yielded the final form of the linguistically annotated corpus.

3.2.2 Extracting Metadata Information

The original TSV per-speech metadata files generated by the ParlaMint script include the following categories:

- **Speech ID.** Unique alphanumeric ID.
- **Title.** The title of the transcription.
- **Date.** Date of speech.
- **Body.** Specifies legislative body, if there are multiple (e.g., in bicameral systems).
- **Term.** Parliamentary term/legislative period.
- **Session.** Terms are subdivided into half-year sessions, usually before and after a summer break.
- **Meeting.** Numbered meetings according to sessions or terms.
- **Sitting.** May mark when a meeting is subdivided into multiple sittings, but this category is empty in ParlaMint-DK and doubles the date in ParlaMint-BG.

- **Agenda.** Possibly meant to include section headings but is empty in both corpora.
- **Subcorpus.** Marks speeches as part of a COVID-19 subcorpus or a reference subcorpus.
- **Speaker role.** Regular, Chairperson or Guest. See part I. in Sections 3.2.3 and 3.2.4 on the different definitions of "Chairperson". There were no "Guest" speakers in either parliament.
- **Speaker MP.** MP or non-MP.
- **Speaker Minister.** Minister or not.
- **Speaker party.** Speaker party abbreviation. In ParlaMint-BG, it was necessary to standardize the party abbreviations, as some were written in Cyrillic.
- **Speaker party name.** Speaker party full name.
- **Party status.** Coalition, Opposition or None.
- **Speaker name.** Full name of speaker.
- **Speaker gender.** Male or female.
- **Speaker birth.** Year of birth.

Having tagged the corpora for sentiment on a by-word basis, it also made sense to add sentiment information to the speech metadata by aggregating over the tokens associated with a speech (see also 4.3.5). The following features were computed for each speech:

- **Sentiment.** The sum of the sentiment scores for all words in the speech.
- **Word total.** The length of the speech.
- **Score.** Sentiment divided by word total, i.e., the speech score. See also section 4.2.4 on scoring procedure.

Not all relevant information in the corpus is equally accessible as annotation or metadata. On multiple occasions throughout the data preparation and analysis process, it became necessary to extend the per-speech metadata with additional extracted features, listed below:

- **Government.** Sitting government at time of speaking. Labels based on the dates given for various governments in the organizational taxonomy provided with each corpus.

- **Caretaker.** Boolean labels indicating whether governments are regular governments or caretaker governments in ParlaMint-BG. Information was obtained from Wikipedia.
- **Bloc.** Left or right (or neither in a minority of cases), based on speaker party. Appendix A reports the parties assigned to each bloc for both parliaments along with their abbreviations in the corpus.
- **Year-Month.** By-month labels generated from existing date labels because the full date is often too granular a unit for analysis.
- **Agenda title.** Section headings indicating the type of proceeding and topic of discussion. In ParlaMint-DK, these were provided as notes in the corpus and were easy to extract. There were no section headings in ParlaMint-BG. In order to obtain them, it was possible to take advantage of a feature of the Bulgarian transcriptions, namely that agenda items, almost exclusively announced by the chairperson, are rendered in all caps. Using the plaintext versions of the speeches and Python's string processing functions, all capitalized sequences over 10 characters long from these speeches were extracted. (Shorter capitalized sequences are usually abbreviations.) By saving the indices of the speeches containing chairperson announcements, all speeches in between such announcements (within the same meeting) could be assigned to the extracted topic title. This procedure resulted in the annotation of 87% of speeches, though some degree of error must be assumed within that.
- **Proceeding type.** Indicates the type of proceeding, i.e., a reading, hearing, negotiation or other. Agenda announcements or section headings, described above, usually begin by naming the type of proceeding or debate, and keywords from these can be used to annotate the speeches. For both corpora, keywords were identified systematically by adding words to a list of proceeding types and then removing all titles containing those words, until all topic titles were covered. In ParlaMint-BG, some titles contained multiple keywords, and in those cases the first one was taken to indicate the true type (e.g., "Decision to vote…"). The keywords, their translations (where necessary) and their corresponding proceeding types are listed in Tab. 3.1. A few very rare types were subsumed under "Other" label – covering 0,1% of speeches in the Danish corpus and about 3% of labelled speeches in the Bulgarian.

- **Committee.** Boolean. During the process of analysis and preparation of the Bulgarian gold standard, it became apparent that there were a large number of speeches, given primarily by coalition members, that announce the work and position of a legislative committee. These speeches have a largely procedural purpose and usually include lengthy quotations from legal and legislative texts. Because this proved significant for the sentiment analysis (see part I. of Section 5.2.2), an attempt was made to distinguish these speeches with a Boolean label. Speeches were labelled based on the occurrence of the phrase "The committee does (not) support…", which, according to observations from annotating an early version of a gold standard, was present in the vast majority of these speeches.

Proceeding type	**Keywords - Denmark**	**Keywords - Bulgaria**
Bill reading	behandling (*treatment*)	четене (*reading*) законопроект (*draft bill*)
Voting	afstemning	гласуване
Questioning	spm. (*question*, abbr.) spørgetime (*question hour*)	изслушване (*hearing*) контрол (*control*) надзор (*supervision*)
Negotiation	forhandling	-
Request	hasteforespørgsel (*urgent request*)	-
Party Leader Debate	partilederdebat	-
Draft Decision	-	проект(и) (*project/draft*, pl.)
Procedural	-	решение (*decision*) избор (*election*) промени (*changes*)
Report	-	отчет
Other	punkt 0 (*item 0*) godkendelse (*approval*) redegørelse (*explanation*) højtidelighed (*ceremony*)	представяне (*presentation*) относно (*concerning*) разисквания (*discussions*) оставка (*resignation*) изпълнение (*enforcement*)

Table 3.1 Types of parliamentary meetings based on keywords from agenda titles. Some proceeding types and keywords do not have analogues in both parliaments.

3.3 Gold Standard Annotation

Another important treatment of the data was the creation of a gold standard for validation and evaluation purposes. 300 speeches or speech excerpts between 10 and 200 (Danish) or 300 (Bulgarian) words were annotated from each national corpus. While the formal cutoff was 100 words, maintaining coherence required keeping the entire segment containing the 100th word. The difference between the Danish and Bulgarian maximum lengths is then due to the difference in the way segments are encoded in the data.

The speeches were sampled randomly and proportionally to party status. Speeches by chairpersons and non-MPs were excluded, as well as Bulgarian speeches marked as "Committee" (see 3.3.2 above). Furthermore, in order to assess annotation reliability, the first 100 speeches from each gold standard were provided to a second native language annotator. The author of this thesis, who is bilingual, was the primary annotator.

Speech excerpts were labelled on an integer scale from -2 to 2, corresponding to the ratings "strongly negative", "mostly negative", "neutral", "mostly positive" and "strongly positive". Annotation guidelines (see Appendix B) instruct annotators to identify emotion based on framing and manner of expression. Emotional language is characterized by the use of emotionally charged words or words with clear positive or negative connotations. It does not include, however, speech acts that can be construed as positive or negative but are expressed neutrally– e.g., statements of support ("We support this proposal.") or critical questions ("How much money will be allocated to this?"). Sentiment is considered separate from content and from the valence of a speech act in itself. This distinction can be justified theoretically, as the position-taking functions of parliamentary speech go beyond expressions of party-level support for particular pieces of legislation (Proksch & Slapin, 2012). For example, an opposition MP may support the contents of a motion but still criticize its timing or wording using negative sentiment.

Annotating the speeches was subject to a large degree of uncertainty. Factors such as quotation, conditionals and complex rhetorical setups made it difficult to evaluate emotional polarity according to the guidelines. Furthermore, a single speech excerpt may contain opposing expressions of sentiment, for example this excerpt from a Danish MP (trans. by Google Translate with the author's corrections):

> ”Thank you very much for a really, really nice speech. I think it is so wonderful that there are more people who dare to stand on the floor of the Parliament and say that this simply requires fundamental changes. We are facing a climate crisis so big that we cannot just do business as usual. To dare to stand on the podium and speak against this growth mentality, which has been booming out there for the past many, many years, and which has indeed caused parts of the climate crisis – I think there is a need for that.

This speech expresses joy and gratitude that an important topic (climate change) is being discussed, but the sentiment toward the issue and its causes ("the growth mentality") is rather negative. As the expressions of positive emotion were more explicit, it was judged by the author as being predominantly positive despite ambiguity. The secondary annotators also reported difficulty in distinguishing sentiment intensity ("mostly" vs. "strongly") and labelled many speeches they were unsure about as neutral. Inter-annotator reliability is addressed in Section 6.1.1.

3.4 Descriptive Statistics

While the gold standard annotation doubled as a type of qualitative data exploration, this section presents a more thorough quantitative introduction of the corpora, highlighting key variables and points of contrast. Additional information and tables can be found in Appendix C and D.

3.2.3 ParlaMint-DK

The ParlaMint-DK 3.0 corpus contains all parliamentary proceedings from October 7, 2014, to July 6, 2022, and spans four full or partial legislative periods and governments (Tab. 3.2). It contains about 40.8 million tokens across 398,610 speeches and 372 speakers. The average speech is 102 tokens long, and 87% of speeches are under 200 tokens long.

Background

The data for the corpus was obtained from the Danish parliament website[3] in XML format (Erjavec et al., 2022). Danish Hansard transcriptions are lightly

[3] https://oda.ft.dk/

edited for readability and syntactic coherence, but are otherwise verbatim (Navarretta & Hansen, 2022). The transcriptions also include section headings, which are incorporated as notes in the corpus. Linguistic annotation was performed with the Text Tensorium workflow management system[4], which includes tools from CST (Center for Language Technology) and Dapipe[5].

Governing period	**Government**	**Coalition**	**Opposition**	**None**
2014-02-03–2015-06-27	Thorning-Schmidt II	S, RV (61)	V, KF, LA, DF (86)	EL, SF (28)
2015-06-28 – 2016-11-27	Rasmussen II	V (34)	S, RV, SF, EL, ALT (85)	DF, LA, KF (56)
2016-11-28 – 2019-06-05	Rasmussen III	V, KF, LA (53)	S, RV, SF, EL, ALT (85)	DF (37)
2019-06-27 – *2022-12-15*	Frederiksen I	S (48)	V, KF, LA, DF, NB (79)	RV, SF, EL, ALT (48)

Table 3.2 Governing periods for cabinets covered fully or partially (italic) by ParlaMint-DK 3.0. Coalition, opposition, and other parties are labelled for each period along with the corresponding number of MPs in parentheses.

Speaker Roles

The distribution of speeches and words across different roles in parliament can be seen in Tab. 3.3. The “Chairperson” category includes everyone in the presidium, i.e., a chairperson and four vice-chairpersons. The “Other non-MP” category appears to consist mostly of temporary substitute MPs.

In general, the number of words spoken a group is roughly proportional to its speaker share. Chairpersons give many short speeches, while Minister MPs disproportionately many words in long speeches. Other non-MPs (substitutes) tend to speak less than their share would indicate, consistent with their provisional status. Breakdowns by party status and gender are available in Appendix C.

[4] https://clarin.dk/clarindk/tools-texton.jsp
[5] https://github.com/ITUnlp/dapipe

Role	Speakers (%)	Speeches (%)	Words (%)	Avg. Speech Length
Chairperson	5,5	52,2	5,8	11,4
Regular MP	73,4	38,2	72,7	194,9
Minister MP	13,1	8,1	18,5	234,4
Minister non-MP	1,2	0,2	0,5	209,7
Other non-MP	7,0	1,3	2,5	191,4

Table 3.3 Distribution of speakers, speeches, words, and speech length by speaker roles in ParlaMint-DK. Speaker percentage is calculated on the basis of total speaker-role pairs.

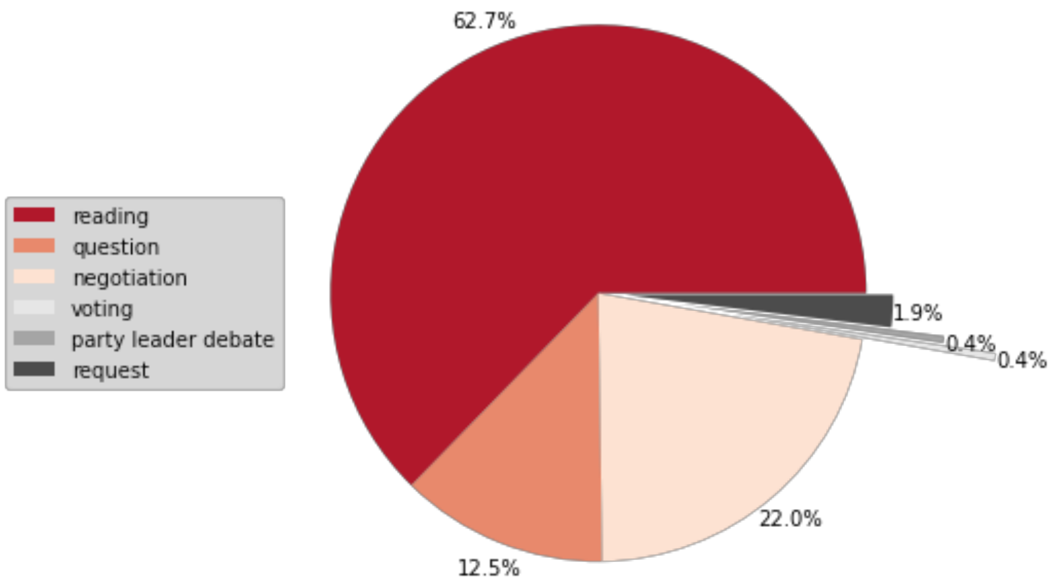

Figure 3.3 Percentage of words spoken during different proceeding types in the Danish parliament 2014-2022.

Types of Proceedings

Proceeding type is a key variable for this investigation and can reveal much about the priorities and function of parliament. Fig 3.3 demonstrates the distribution of speech among different parliamentary functions. Most words are spoken during readings of legislation, followed by negotiations (Dan. *forhandlinger*, more general thematic debates) and questions (ministerial hearings). The miniscule percentage of other speech types (e.g., commemorative) is not included in the graph. The pattern that emerges is a focus on substantive legislative matters, with very little procedural speech (e.g., announcements for voting).

3.2.4 ParlaMint-BG

The ParlaMint-BG 3.0 corpus contains all parliamentary proceedings from October 27, 2014, through July 29, 2022. This period spans three terms and seven governments, four of which are interim caretaker governments, reflecting the political instability in Bulgaria during this time (Tab. 3.4).

The corpus contains 26,4 million tokens across 210 017 speeches and 849 speakers. On average, speeches are 126 tokens long, although they range from a single token to 8908 on the maximum end. However, only 1,3% of speeches are longer than 1000 tokens, and 80% are under 200 tokens long.

Background

The Bulgarian ParlaMint team obtained the texts of plenary sessions from the official website of the Bulgarian National Assembly[6], while metadata information was gathered from both the parliament website and sources such as Wikipedia (Erjavec et al., 2022). Linguistic annotation was performed with the CLASSLA pipeline (Ljubešić & Dobrovoljc, 2019). A large number of "incidents" – shouts, laughter, clapping or other relevant happenings – are encoded as notes in the data. Section headings, on the other hand, are not present in the original transcriptions or the corpus.

Governing Period	Government	Coalition	Opposition	None
2014-08-06 – 2014-11-07	Bliznashki (C)	-	-	All
2014-11-07 – 2017-01-27	Borisov II	GERB, ABV*, RB, (126)	BSPLB, MRF, PF (89)	BGUn, AP, ABV* (35)
2017-01-27 – 2017-05-04	Gerdzhikov (C)	-	-	All
2017-05-04 – 2021-05-12	Borisov III	GERB, UP (122)	BSPFB, MRF, VOLYA (118)	Independent (17)

[6] https://www.parliament.bg/bg/plenaryst

2021-05-12 – 2021-09-16	Yanev I (C)	-	-	All
2021-09-16 – 2021-12-13	Yanev II (C)	-	-	All
2021-12-13 – *2022-08-02*	Petkov	WCC, DB, BSPFB, TISP (134)	GERB-UDF, RP, MRF (106)	Independent (7)

*In coalition until 2016-05-18

Table 3.4 Governing periods for cabinets covered fully or partially (italic) by ParlaMint-BG 3.0. Coalition, opposition, and other parties are labelled for each period along with the total number of MPs in parentheses. (C) indicates a caretaker government.

Speaker Roles

Tab. 3.5 presents the participation of various groups according to their speaker roles. In the Bulgarian data, "Chairperson" refers to both parliament chairpersons and prime ministers, as well as their respective deputies. These can however be distinguished, as chairpersons are chosen from among MPs, while prime ministers are not. The difference is starkly apparent with prime ministers giving speeches almost 10 times longer than chairpersons, though chairpersons speak much more frequently.

Role	Speakers (%)	Speeches (%)	Words (%)	Avg. Speech Length
Chairperson MP	3,8	48,0	13,5	35,4
Chairperson non-MP*	2,0	0,8	2,1	313,9
Regular MP	67,2	45,5	71,5	194,9
Minister non-MP	2,1	0,4	1,2	209,7
Other non-MP	24,6	5,1	11,6	191,4

*Prime Minister and deputies

Table 3.5 Distribution of speakers, speeches, words, and speech length by speaker roles in ParlaMint-BG. Speaker percentage is calculated on the basis of total speaker-role pairs.

Ministers also give longer speeches than regular MPs, though very infrequently, as they are not MPs and only appear during hearings. Almost a quarter of speakers fall into the ill-defined "Other non-MP" category. This category appears to include guest speakers of various sorts, including committee members, vice-ministers, and anonymous speakers, as well as some mistakes[7]. “Other non-MP” speakers are rarely party members and give a small number of speeches, although they represent almost a quarter of speakers.

Proceeding Type

Fig. 3.4 shows the proportion of words spoken within each proceeding type, a more heterogenous group in the Bulgarian parliament. Debates on legislation appear to happen both during readings and during voting meetings, and these two categories make up just under half of all speech. The biggest single category is “draft decisions” (Bul.: *проект за решение*), during which MPs discuss and vote on various procedural and executive decisions, including the creation, makeup and tasks of committees, changes to the parliamentary rulebook, approval of funding, etc. Since anyone can propose draft decisions, this appears to be one of the main strategies for power jockeying in the parliament, based on a limited manual inspection of speeches.

Ministerial questioning and parliamentary control take up relatively few words. Procedural speech includes other final decisions, nominations, and changes to commission memberships, and could potentially be grouped with draft decisions.

Caretaker Governments

Taken together, caretaker governments cover almost a year of the corpus, but the debates during such governments only make up a tiny proportion of parliamentary speech – 5,3% of MP speeches and 4,2% of MP words. As much of the regular functioning of commissions and parliament is paused, very little legislative debate takes place under such governments, with hearings (33% of speeches) and draft decisions (30%) predominating instead.

[7] For example, Kiril Petkov is neither labelled as minister (of the economy) during the Yanev I government, nor as a “Chairperson” while he was prime minister, which means he incorrectly ends up in the “Other non-MP” category both times.

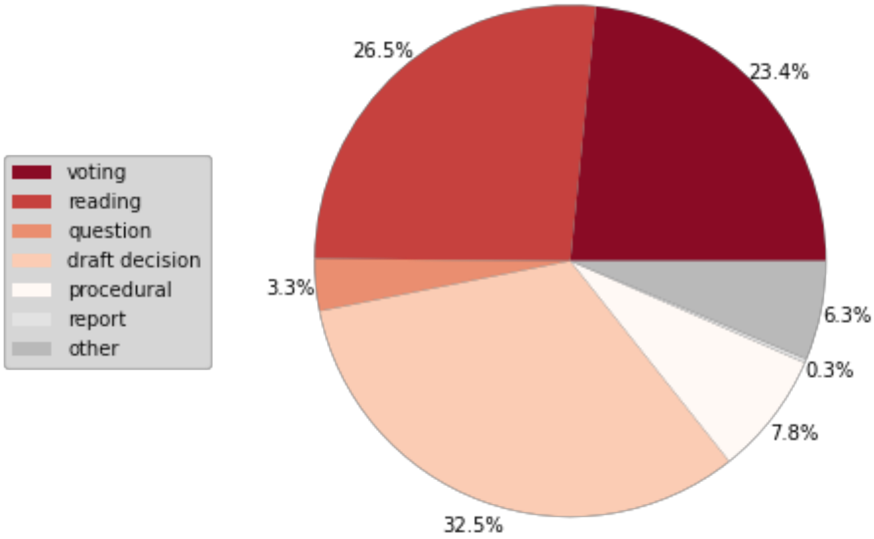

Figure 3.4 Proportion of words spoken during different proceeding types in the Bulgarian parliament 2014-2022.

3.2.5 Comparative Summary

The differences between the two parliaments can be summarized in a few points. Ministers are disproportionately responsible for speech in Denmark, but not in Bulgaria, where they are not MPs. Bulgarian chairpersons give lengthier speeches than Danish ones, which may reflect the greater quantity of procedural speech in the Bulgarian parliament in general. Legislative debate makes up less than half of all speech in the Bulgarian National Assembly, with procedural and committee matters taking a much more prominent place than in the Danish parliament, which is focused on legislature and policy debates.

4 METHOD

This chapter presents the theory and implementation of the methodology. Section 4.1 provides the background and motivation for the chosen approach to sentiment analysis, while Section 4.2 lays out the NLP theory and resources behind the methodology. Finally, Section 4.3 details the practical process of implementation.

4.1 Emotional Polarity

Sentiment analysis (or opinion mining) encompasses a large range of textual techniques for quantifying all kinds of subjective states – opinions, attitudes, emotions, etc. Emotional polarity is perhaps the first and simplest of constructs studied by the field: The identification of "negative" and "positive" emotion in texts such as movie reviews or customer feedback fueled its initial development (Medhat et al., 2014), though it has since expanded its scope to analyze texts from social media, political speech, journalistic articles etc. (e.g., Hutto & Gilbert, 2014; Young & Soroka, 2012).

Among the main advantages of sentiment analysis are the reliability and scalability afforded by automation; challenges lie in insufficient validity and the strong context-dependency of linguistic markers of sentiment, which can be difficult to capture. Sarcasm, tone, and register interact in complex ways with the "basic" meanings of words, and even human judges can struggle to consistently classify words or utterances by sentiment (Abercrombie, 2021). Nevertheless, many sentiment analysis techniques have demonstrated their utility and achieved remarkable results across various domains.

4.1.1 Sentiment Lexica

Approaches to measuring emotional polarity fall into two main categories: lexicon-based and machine learning. Human-annotated psycholinguistic lexica have a long provenance and history of use in various fields, and these different "backgrounds" are reflected in the intentions, scope, and methodologies of the lexica. The General Inquirer (Stone et al., 1966) and the Lexicoder Sentiment Dictionary (LSD) (Young & Soroka, 2012) were created by compiling and filtering existing word lists and dictionaries. The Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001) and Affective Norms of English Words (ANEW) (Bradley & Lang, 1999) were developed based on aggregations of human judge ratings. VADER (Hutto & Gilbert, 2014) compiles and extends other sentiment dictionaries and incorporates intensity information on both a lexical and syntactic level for a more context-sensitive approach.

All of the aforementioned lexica were developed for English. Sentiment analysis for other languages often relies on adaptation/translation of English lexica or machine translation. AFINN (Nielsen, 2018) is a Danish adaptation of ANEW consisting of 3552 words manually annotated on a scale from -5 to +5 by the author. More recently, Lauridsen et al. (2019) have developed SENTIDA for the most frequent Danish words via manual annotation by three judges combined with heuristics to account for negation, exclamation, and intensifying modifiers. The Danish Sentiment Lexicon (Nimb et al., 2022) achieves high lexical coverage by using a comprehensive thesaurus resource.

Bulgarian sentiment lexica appear to be limited to Kapukaranov & Nakov (2015), who generate their lexicon automatically using point-wise mutual information (PMI) on a dataset of positive and negative movie reviews, making it more of a combination method (see below). It is fairly domain-specific and not validated on other data.

4.1.2 Machine Learning for Sentiment Analysis

Creating and validating sentiment lexica is time-consuming and labor-intensive, but once created, they are computationally simple to implement. Parts of this process can be automated by training machine learning algorithms such as Naïve Bayes or Support Vector Machines to learn the textual features which characterize positive and negative emotion. More recently, deep models such BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2018) have been leveraged to capture even more subtle features and linguistic relationships relevant to sentiment (Tang et al., 2016). The main drawback of these approaches is the relative paucity of labelled training data, which is necessary but usually limited to certain genres, domains, and languages. Furthermore, training can become quite computationally expensive.

Combining machine learning and lexicon methods is one way to mitigate this factor and speed up the lexicon creation process, allowing for an unsupervised analysis of unlabeled data. This is usually done by starting with a set of known positive and negative words and computing measures of co-occurrence strength to expand the list, e.g., using point-wise mutual information (Kapukaranov & Nakov, 2015; Turney & Littman, 2003). Rheault et al. (2016) leverage co-occurrence through word embeddings by expanding a hand-validated list of sentiment words. Their approach is elaborated below and forms the methodological background of this thesis.

4.2 Multi-Lingual Sentiment Lexica with GloVe

Rheault et al. (2016) develop their approach with a concern for domain-specificity when analyzing the Hansard of the British House of Commons. Many sentiment lexica do not transfer well across domains (Meden, 2022) for reasons of limited vocabulary or bias. The solution, then, is to start with a conceptually neutral seed list of positive and negative words (e.g., *good/bad*). Starting from these words and sing recursive dictionary search, Rheault et al. (2016) compile a list of terms with unambiguous affective orientation and no political or procedural meanings, keeping the 200 most frequently occurring words in each category as the final seed. The entire corpus of parliamentary speech is then embedded as a vector space model using GloVe (Pennington et al., 2014). The words closest in vector space to the positive and negative seeds are added to the lists, expanding them considerably with domain-specific terms. This final expanded dictionary is used to score the speeches and compare different time periods and party groupings.

Apart from domain-specificity, this approach has the advantage of being largely language-independent, relying on simple resources like frequency lists and PoS-taggers. Multi-lingual neural models are very dependent on the source and quality of data from different languages, and fine-tuning is computationally expensive. Lexicon translation is another low-tech solution. Proksch et al. (2019), a promising study in that vein, apply machine translated versions of the Lexicoder Sentiment Dictionary (Young & Soroka, 2012) to parallel corpora from the European Parliament. They report a strong correlation between dictionary and hand-coded scores for parallel sentences across most languages in their study, with the exception of Slovenian, Greek and Bulgarian. The poor performance for Bulgarian is the main reason this translation approach was rejected for this project. Translating a German version of the LSD also yielded much worse results across the board, demonstrating the potential volatility of translation.

The chosen approach from Rheault et al. (2016) will hopefully produce more valid measures, particularly for Bulgarian, and some degree of cross-national comparability. Its adaptability to smaller and less researched languages and parliaments will also be an interesting avenue for replication.

4.2.1 WordNets

The first stage of compiling seed dictionaries in Rheault et al. (2016) relies largely on WordNet 3.0 (Fellbaum, 2010). Wordnets are lexical databases that encode semantic relationships between words, primarily synonyms but also hyper/hyponyms (super- and sub-categories), meronyms (part-whole relations) and others. Synonyms are organized into synsets with definitions and usage examples. WordNets have become crucial resources for NLP tasks such as word-sense disambiguation and information retrieval and have been developed for many different languages.

The Danish DanNet (Pedersen & Asmussen, 2006) takes as its main lexical resource the corpus-based dictionary Den Danske Ordbog (DDO: Hjorth & Kristiansen, 2005). Since the recursive search process for creating seed dictionaries was manual, it was actually easier to directly use DDO and its function "Ord i nærheden" (Words nearby) to collect synonyms. The DanNet website[8] emphasizes hyponymy and ontological descriptions to a greater degree, making sysnet connections less visible.

The Bulgarian BulNet was built as an expansion of the original Princeton WordNet using machine translation and various Bulgarian lexical resources (Koeva, 2021), so it is richly annotated with semantic, derivational and domain information[9]. For the purpose of this study, only the basic synset information was used.

4.2.2 Word Embeddings

[8] https://andreord.nors.ku.dk/

[9] http://dcl.bas.bg/bulnet/

To achieve domain-specificity by expanding the initial seed lists, it is necessary to compute a measure of semantic similarity. This is where word embeddings come in – as vectorized representations of words, they are encoded so that semantically similar words are closer in vector space than semantically dissimilar words, making distance a proxy for similarity. Creating word embeddings is an unsupervised machine learning task that leverages word co-occurrence and context statistics with various available approaches.

Following Rheault et al. (2016), GloVe (Global Vectors) embeddings (Pennington et al., 2014) were chosen for this task. While context-sensitive embeddings generated by language models like ELMo (Peters et al., 2018) can disambiguate various word senses and provide more accurate semantic encodings, they require greater computing power and need to be used for inference as well as training. By contrast, once trained, "bag-of-words" embedding algorithms like GloVe create a single vocabulary of embeddings that is easily accessed. And since results from the embeddings are going to be applied to the same dataset they were generated from, it is not a problem that GloVe cannot deal with out-of-vocabulary words.

4.2.3 The GloVe Algorithm

GloVe was originally developed as an open-source project at Stanford University, and the original algorithm and implementation have been publicly available on GitHub[10] since its inception. The method combines advantages from two other embedding methods – matrix factorization and shallow context windows – into a global log bilinear regression model. While it starts by building a word co-occurrence matrix based on global vocabulary, it also defines a weighted context window for co-occurrence, so that words closer to each other within the context window are weighted higher than those further apart. This step allows GloVe to capture both global statistics and local context.

The creation of the weighted co-occurrence matrix is the most computationally expensive part of the process. Once generated, the rows or columns of the matrix can be considered as word or context vectors. The training objective of GloVe is then formulated as a least-squares problem of minimizing the difference between the dot product of a word vector (w_i) and context vector (w_j) and the logarithm of their co-occurrence count X_{ij} (along with bias terms):

[10] https://github.com/stanfordnlp/GloVe

$$\min \sum_{i,j=1}^{V} \left(\mathrm{w}_i^T w_j + b_j + b_i - \log(X_{ij})\right)^2$$

As X_{ij} is an estimate of the words' probability of co-occurrence, taking its logarithm will associate ratios of co-occurrence probabilities, which can encode meaning, with vector differences in the word vector space. This is possible because the logarithm of a ratio is equal to the difference of logarithms.

Finally, a weighting function is applied in order to ensure that very rare or very frequent co-occurrence values are not overweighted in training:

$$f(X_{ij}) = \begin{cases} \left(\frac{X_{ij}}{x_{max}}\right)^a & \text{if } X_{ij} < x_{max} \\ 1 & \text{otherwise} \end{cases}$$

With $x_{max} = 100$ and $a = 0.75$ being the default settings of the function parameters. The formulation of the cost function J is therefore:

$$J = \sum_{i,j=1}^{V} f(X_{ij})\left(w_i^T w_j + b_j + b_i - \log(X_{ij})\right)^2$$

The cost function is minimized using stochastic gradient descent (SGD), a machine learning technique that calculates cost function gradients and moves parameters in a minimizing direction. For the function J, both word vectors and biases are minimized simultaneously, yielding four gradients:

$$\frac{\partial J}{\partial w_i} = \sum_{i,j=1}^{V} 2 \cdot f(X_{ij}) w_j \left(w_i^T w_j + b_j + b_i - \log(X_{ij})\right)$$
$$\frac{\partial J}{\partial w_j} = \sum_{i,j=1}^{V} 2 \cdot f(X_{ij}) w_i \left(w_i^T w_j + b_j + b_i - \log(X_{ij})\right)$$
$$\frac{\partial J}{\partial b_i} = \sum_{i,j=1}^{V} f(X_{ij})\left(w_i^T w_j + b_j + b_i - \log(X_{ij})\right)$$
$$\frac{\partial J}{\partial b_j} = \sum_{i,j=1}^{V} f(X_{ij})\left(w_i^T w_j + b_j + b_i - \log(X_{ij})\right)$$

At each step in the learning process, the parameters are changed in the minimizing direction according to the learning rate α with a step function:

$$x_{t+1} = x_t - \alpha g_t,$$

Where x_t is a weight at iteration t, α is the learning rate, and g_t is the gradient of x_t. The authors of the GloVe paper and algorithm implementation use a modification of SGD called AdaGrad (Duchi et al., 2011), which adapts individual learning rates for each weight by tracking the history of gradients, obviating the need for step size tuning. The step function for AdaGrad is given by:

$$x_{t+1,i} = x_{t,i} - \frac{\alpha}{\sqrt{\sum_{\tau=1}^{t-1} g_{\tau,i}^2}} g_{t,i}$$

Where t is the iteration step and i is the weight component. The embedding weights are updated for a set number of iterations or until some convergence criterion is met.

4.2.4 Sentiment Scoring

Once embeddings have been created for the whole vocabulary, the initial seed list of sentiment words can be expanded. This requires that all seed words are actually present in the corpus vocabulary and frequent enough to generate meaningful embeddings.

When comparing the corpus vocabulary to the seed words, semantic nearness is defined as cosine similarity. Thus, the sentiment score s_i for a word is given by the difference between the summation of its cosine similarity to a set of positive seed words P and to set of negative seed words Q (Rheault et al., 2016):

$$s_i = \sum_{p=1}^{P} \frac{\mathbf{w}_i \cdot \mathbf{w}_p}{\|\mathbf{w}_i\| \|\mathbf{w}_p\|} - \sum_{q=1}^{Q} \frac{\mathbf{w}_i \cdot \mathbf{w}_p}{\|\mathbf{w}_i\| \|\mathbf{w}_p\|}$$

Where w_i is the embedding for word i. The scores are normalized, and the words with the most and least positive scores are then used to expand the seed dictionaries and create the final domain-specific sentiment lexica.

When applying sentiment analysis to text, sentiment word scores are aggregated across the desired time periods/utterances/etc. and normalized by the total number of words spoken. Furthermore, in order to account for the effect of negation, an extra parameter θ is multiplied to the score. θ is equal to 1, except after a negation and until the next punctuation mark, when it is 0. This means that negated sentiment words will not be counted towards the final score. This yields the formula:

$$y_t = \frac{\sum_{i=1}^{n_t} \mathbf{1}\{w_{it} \in L\} s_i \theta_{it}}{\sum_{i=1}^{n_t} \mathbf{1}\{w_{it} \in L\}}$$

Where n_t is the total number of words for the period t, and L is the corpus lemma vocabulary. Higher values of y_t indicate more positive debates. Rheault et al., (2016) point out that, since positive words are used more frequently in speech, the average score is likely to be positive, so scores will not distribute normally around 0. This means that relative, rather than absolute, differences in sentiment are of interest.

The evaluation and validation procedures for the sentiment analysis results are described in Section 6.1.

4.3 Implementation

4.3.1 Seed Dictionaries

Following Rheault et al. (2016), the initial seed list was produced manually using recursive search of in BulNet and DDO, starting with the most basic positive and negative words as initial "poles" (Tab. 4.1) and working out branching synset connections.

Language	Positive			Negative		
English	good	love (n./v.)	happy	bad	hate (n./v.)	sad
Danish	god	elske/ kærlighed	glad	dårlig	hade/had	trist
Bulgarian	добър	обичам/ любов	щастлив	лош	мразя/омраза	тъжен

Table 4.1 Core sentiment words for English (Rheault et al., 2016), Danish and Bulgarian.

The resulting lists were compared across languages (including the English list provided by Rheault et al., 2016) to make sure there were no obvious missing terms. This manual task could have been automated to a greater degree, but it did not require that much work and allowed for evaluation of words on the go. The inclusion criteria for seed lemmas were taken directly from the supplementary material to Rheault et al. (2016), quoted below:

> 1. Polysemous seed words need to have an unambiguous emotional orientation, which means that multiple meanings of the same word used as the same PoS must not have opposite polarities.
> 2. Seed words cannot be the name of an institution, parliamentary procedure, or political topic (excluded are words such as *war*, *dispute*, *unemployment*, and so forth).
> 3. Seed words need to be basic and common words of everyday language.

An issue with the first criterion is that the frequency lists used to rank the seed lemmas for both languages later on did not include part-of-speech tagging, so words with inconsistent polarity across PoS were excluded. More examples of excluded words and explanations for both languages are listed in Appendix F. The final positive lists numbered 285 words for Bulgarian and 336 for Danish, while the negative lists were longer, 336 and 559 words for each language respectively.

Seed words were PoS-tagged and lemmatized to get rid of duplicates and then ranked by frequency. For maximally congruent results, the same processing software used by the ParlaMint researchers was chosen: CST tools[11] for Danish and the CLASSLA library (Ljubešić & Dobrovoljc, 2019) for Bulgarian. The annotations were manually corrected where necessary. Since the seed dictionaries represent domain-neutral sentiment terms, their frequency was ranked according to external sources. For Danish, this was a frequency list compiled by Jørg Asmussen, containing the 30 thousand most frequent lemmas in the DSL corpus[12]. The Bulgarian list, provided by BulTreeBank[13] (Simov et al., 2002), lists the 100 thousand most frequent tokens in a reference corpus. The BTB list had to be lemmatized and its frequencies updated for compatibility with the seed lists. Based on these two resources, the 200 most frequent positive and negative lemmas were kept as seeds along with their PoS-tags.

4.3.2 GloVe Embeddings

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

[12] https://korpus.dsl.dk/resources/licences/dsl-open.html#en

[13] https://bultreebank.org/bg/resources/

The two corpora were embedded with Glove using the Stanford NLP source code[14] in C. The vector space model has 150 dimensions, a window size of 15, a minimum frequency of 5, and otherwise uses default parameters. Vectors initially had 300 dimensions following Rheault et al. (2016), but a qualitative evaluation of the matched words indicated that halving the size yielded more relevant matches.

The corpora were embedded based on the already tokenized and lemmatized TEI files. In order to disambiguate homophonous tokens with different parts of speech, lemmas were concatenated with their PoS-tags before embedding (e.g., "tale-VERB"), following Rheault et al. (2016). This partially alleviates the problem of polysemy.

With these specifications, the Danish corpus yielded a vocabulary of 56 210 lemma embeddings, while the Bulgarian corpus yielded 35 162 embeddings, reflective of their differences in size. More than half of unique lemmas in both corpora appeared fewer than 5 times and were excluded.

4.3.3 Lexicon Expansion

The first step in expanding the seed lemma lists was making sure all lemmas were present in the embedding vocabulary. Missing lemmas were excluded, and the lists were padded with lower-frequency seed lemmas to reach the original 200 count for each sentiment category. The embeddings for positive and negative seed lemmas were selected and aggregated into arrays. Similarity scores were then computed for each of the remaining vocabulary lemmas, before being normalized and ranked.

Numerals, proper nouns, and punctuation were filtered out of the ranked list, and it was also necessary to impose further frequency limits. Visual inspection of the matches, particularly in the Bulgarian positive list, revealed that very frequent and very infrequent words tended to be less relevant, i.e., function words or common verbs. To mitigate this issue, words outside a frequency interval of 100 to 15 000 were removed. The top 10 matched words for each language and sentiment category are listed in Appendix F with translation.

The 500 most and least positive expansion lemmas were then added to the original seed lemmas along with their similarity scores, which are used as weights. In effect, this privileges the original seed words, which have full weights of ±1.

[14] https://github.com/stanfordnlp/GloVe

Rheault et al. (2016) include 1000 lemmas from each category, but match quality clearly deteriorated around the 500th mark, likely due to smaller corpus size. Combining seed and expansion lemmas yielded ParLex BG and ParLex DK, sentiment dictionaries of 700 positive and 700 negative lemmas each.

4.3.4 Sentiment Tagging

With the sentiment lexica in place, the last step was tagging the corpora. Each annotated CSV file was loaded, and lemmas were tagged by lookup in ParLex, adding a new column. Words not in the sentiment lexicon were tagged as 0, as well as words between negation words[15] and punctuations marks.

The metadata information for each speech was also annotated with a raw sentiment score, word count, and normalized sentiment score, allowing for both by-speech and broader levels of analysis.

Finally, the gold standard speech excerpts were scored based on the sentiment-annotated CSV files.

[15] Danish: ikke, hverken, ingen, aldrig
Bulgarian: не, нито, никой, никога

5 ANALYSIS

This chapter presents the results of the sentiment analysis. Section 5.1 describes the vocabulary and themes present in positive and negative speech for both corpora. Section 5.2 analyzes differences in sentiment according to party status, gender, and proceeding type, and investigates the temporal development of sentiment in the Danish and Bulgarian parliaments. Findings are summarized and compared in Section 5.3.

5.1 Qualitative Keyword Analysis

This section aims to characterize the sentiment vocabulary of both parliaments. The thematic analysis is based on a qualitative review of the most frequently used sentiment words in both categories of emotional polarity. In the sections below, "theme" refers to general semantic patterns among words, while "topics" are more concrete themes relating to political subjects (e.g., economy or immigration).

5.1.1 Danish Keywords

According to analysis with ParLex, there are about 2,1 million sentiment words in the ParlaMint-DK corpus, comprising 5,3% of total words. Common sentiment words are overwhelmingly positive. In fact, out of the 100 most frequent sentiment words in the corpus, only five are negative, and among all sentiment tokens found in the corpus, a mere 15,8% are negative. This is in line with findings from Rheault et al. (2016) and helps explain the fact that the average sentiment rating for speeches is positive (see Section 5.2). Nevertheless, the lexicon is likely overestimating positive sentiment due to a number of highly frequent function words other possibly irrelevant terms added during sentiment expansion (see also discussion in 6.2.1).

Notable themes from the 100 most frequent positive keywords are listed in Tab. 5.1. Positively charged speech in the Danish parliament is commonly characterized by themes such as security, growth, openness, and partnership. MPs speak forcefully, using emphasis and focusing on the necessary action and initiative needed to achieve their goals. The temporal orientation of positive speech is continual, pointing toward the future. Political terms such as parties or positions tend to be mentioned in a positive light, possibly because they are the agents of desired change.

Positive Words	
Emphasis	vigtig, afgørende, understrege, prioritere, særlig *important, crucial, emphasize, prioritize, special/especially*
Security	sikre, styrke, passe, sørge, bakke (op), hjælpe, bidrage *secure, strengthen, look after, provide for, back up, help, contribute*
Partnership	samarbejde, medarbejder, hinanden, fællesskab, fælles, dialog *collaboration, co-worker, each other, community, joint, dialogue*
Action	indsats, løsning, initiativ, ordning, idé *effort, solution, initiative, arrangement, idea*
Political	konservativ, socialdemokrat, europæisk, formand *conservative, social democrat, European, chairman*
Growth	bygge, udvikle, starte *build, develop, begin*
Temporal	fremtid, fortsat, samtidig, starte, altid *future, continually, simultaneously, begin, always*
Openness	adgang, åben, fri *access, open, free*

Table 5.1 Thematic groupings identified among the 150 most common positive sentiment words in ParlaMint-DK.

Tab. 5.2. presents the results of a similar exercise performed for the most frequent negative words (though these are less frequent overall, as mentioned above). In general, negative words are less subtle – more concrete and intense, their negativity being explicit rather than contextual. Among the more abstract themes are causality and loss. When MPs speak negatively, they emphasize the extent, causes and intensity of an issue – presumably in order to justify a particular solution. Themes of danger and harm are obvious, but there is also evidence of specific topics. Environmental and economic issues are mentioned, but the most prominent topics relate to multiculturalism and diversity: on the one hand, issues related to refugees and integration, and on the other, problems of discrimination and racism. These two different framings are consistent with the prominence and polarization this topic has engendered in Danish politics (Holm, 2006; Navarretta et al., 2022).

Negative Words	
Causality	begå, udsætte, medføre, opstå, indgreb, følge, påvirkning *commit, expose, cause, emerge, intervention, follow, effect*
Danger	risiko, risikere, potentielt, farlig, frygte, true *risk, risk (v.), potentially, dangerous, fear, threaten*
Harm	skade, vold, overgreb, drab, krænkelse, chikane *injury, violence, assault, murder, violation, harassment*
Loss	mangle, miste, tab, savne *lack, lose, loss, miss*
Intensity	dyb, alvorlig, ekstrem, total *deep, severe, extreme, total*
Conflict	uenighed, misforståelse, kritisere, anklage *disagreement, misunderstanding, criticize, accuse*
Hate	diskrimination, racisme, hadforbrydelse, forskelsbehandling *discrimination, racism, hate crime, discrimination*
Immigration	ghetto, terrorangreb, fremmed, flygtningestrøm *ghetto, terrorist attack, alien, refugee flow*
Climate	klimaforandring, forurening *climate change, pollution*
Economy	konkurs, fyre, finanskrise *bankruptcy, fire (v.), financial crisis*

Table 5.2 Thematic groupings identified among the 150 most common negative sentiment words in ParlaMint-DK.

5.1.2 Bulgarian Keywords

ParlaMint-BG contains 2.06 million sentiment words, making up 7,8% of the total. Interestingly, the two most frequent sentiment words in this corpus are negative ("problem" and "against"), although positive words dominate overall, comprising 95 out of the 100 most frequent sentiment words. Of all sentiment tokens used in the corpus, only 15,2% are negative, comparable to the results from the Danish corpus. Highly frequent function words and other seemingly irrelevant terms among are even more prominent among the expanded positive keywords for Bulgarian.

Nevertheless, it is possible to make out themes among the positive words (Tab. 6.3). Alongside adjectives of emphasis, there is a category of adverbs expressing accordance and consequence. This may reflect that, in positive speeches, MPs will want to show that their positions are consistent with logical and legal principles. Politeness terms have also made it into the list of frequent positive keywords, particularly ones referring to women. When MPs speak to or about their female colleagues, they may moderate their speech in an expression of benevolent sexism (Glick & Fiske, 1997; Haselmayer et al., 2022).

Procedural terms referring to elections and parliament also appear to be mentioned in positive contexts, indicating trust in the democratic process. Dative terms, referring to giving and taking, are very prominent on the list for unclear reasons. It is possible that they are used in connection with requests and proposals on the part of MPs, or like function words, they may appear due to issues with embedding quality.

Positive Words	
Emphasis	важен, необходим, сериозен *important, necessary, serious*
Consequence	съгласно, съответно, следва *according to, respectively, follows*
Politeness	госпож, дама, госпожа *Mrs. (voc.), lady, Mrs.*
Democracy	парламент, избор, дебат, кандидат *parliament, election/choice, debate, candidate*
Dative	взема, давам, получа, внеса *take, give, receive, bring in*
Temporal	започна, създаване, продължа *start, creation, continue*
Economy/Finance	финанси, икономика, финансиране, оценка *finances, economy, financing, valuation*
Security	подкрепа, защита, отбрана, гарантирам *support, protection, defense, guarantee*
Partnership	участие, участвам, заедно *participation, participate, together*

Table 5.3 Thematic groupings identified among the 150 most common positive sentiment words in ParlaMint-BG.

A summary of the negative themes among the most frequent keywords is shown in Tab. 5.4. Straightforwardly negative themes such as harm and danger dominate, but more abstract ones like causality and declaration are also present.. While specific negative topics tend to be low on the frequency list, the most frequent of them contains words related to corruption and embezzlement scandals, which are commonplace in Bulgarian politics (Tanasoiu & Racovita, 2012).

Negative Words	
Causality	заради, последица, предизвиквам, последствие, в следствие *because of, consequence, provoke, outcome, as a result*
Intensity	тежък, краен, масов, системен *severe, extreme, massive, systemic*
Loss	липса, загуба, пропуск, недостиг, отстраняване *lack, loss, omission, deficiency, removal*
Harm	инцидент, вреда, щета, бедствие *incident, harm, damage, calamity*
Danger	опасност, заплаха, опасен, страх, рисков *danger, threat, dangerous, fear, risky*
Conflict	конфликт, спор, напрежение, противоречие, критика *conflict, dispute, tension, controversy, criticism*
Declaration	констатирам, твърдение, установявам, изразявам *assert, statement, establish, express*
Corruption	корупционен, скандал, манипулация, злоупотреба *corrupt, scandal, manipulation, misappropriation/misuse*

Table 5.4 Thematic groupings identified among the 150 most common negative sentiment words in ParlaMint-BG.

5.1.3 Keywords: Comparative Summary

A qualitative sentiment keyword analysis of both corpora reveals both similarities and differences between the national parliaments. While Bulgarian MPs tend to use slightly more sentiment words, both nationalities use very similar proportions of positive to negative words. Positive keywords tend to be more abstract, with themes such as security, partnership, and emphasis shared in common between both parliaments.

As to negative keywords, common themes are danger, conflict, harm, and loss, as well as adjectives related to scope and intensity. The Danish themes contain multiple concrete topics relating to questions of immigration, discrimination, and environmentalism, while the only reasonably frequent topic on the Bulgarian list pertains to corruption, reflecting something of the political priorities in each country. Surprisingly, while keywords relating to finance and economics figure in the negative list for Denmark, they are positive in the Bulgarian parliament.

5.2 Sentiment Analysis

The speeches considered for sentiment analysis cover all proceeding types and, unless stated otherwise, are spoken by regular MPs. Speeches from Faroese and Greenlander MPs are however excluded. As the sentiment measure is positively biased, all aggregates are expected to be positive, so the quantity of interest is the relative difference in sentiment across various groups, rather than their absolute values. Sentiment scores for various groups (e.g., "coalition men") are computed by treating all speeches within the category as a single document, summing sentiment words cores and norming by document length (see Section 4.2.4).

The analysis was performed by writing all per-speech metadata (including extracted features) into one master file which could be then loaded and manipulated as a Pandas (McKinney, 2010) DataFrame. Plots have been created with Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021).

5.2.1 The Danish Corpus

I. Sentiment and Party Status

When computing sentiment scores by party status, speakers have been further subdivided into male and female, as prior work indicates women may express more positive emotion (Haselmayer et al., 2022). Minister status is also considered for coalition speakers.

Fig. 5.1 shows a clear sentiment divergence between coalition speakers and others. Within a coalition, ministers express slightly more positive sentiment than regular MPs on average. The difference is small but consistent with prior work, which attributes this to ministers' status as negotiators with the opposition (Rudkowsky et al., 2018). Opposition and other MPs group together at a lower relative sentiment. This is expected for the opposition, which is likely to be a critical voice in parliament, but it is surprising for "third-party" speakers, who are usually considered to be supporting actors toward the coalition (Green-Pedersen

& Skjæveland, 2020). At least with regards to speech and sentiment, these results indicate that they may actually function more similarly to the opposition.

In terms of gender, there is a consistent positive effect on sentiment for women speakers, across all party groups, though the difference is most pronounced for coalition MPs. However, the effect of gender is much smaller than that of party status, the patterns of which persist across male and female speakers, indicating that speakers' function in parliament overrides any gendered roles.

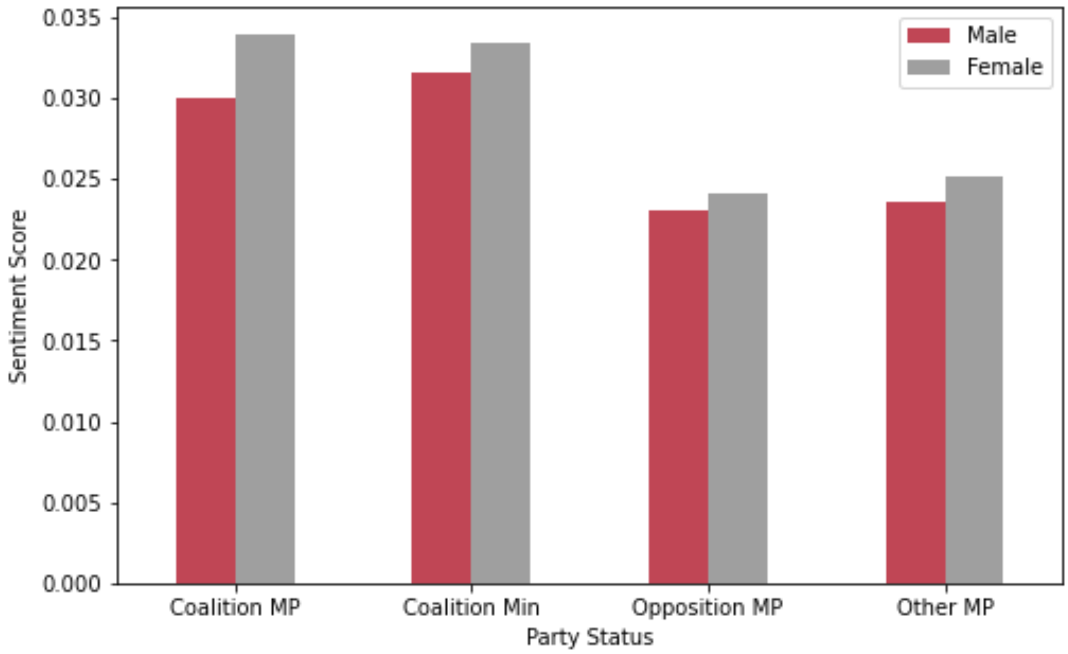

Figure 5.1 Sentiment by party status and gender in ParlaMint-DK. Coalition speakers are further subdivided into regular MPs and Ministers.

In order to better understand what is driving the divergence between coalition and opposition sentiment, scores were also computed on a by-speech basis. The histogram in Fig. 5.2 shows a probability distribution of scores for coalition and opposition speeches.

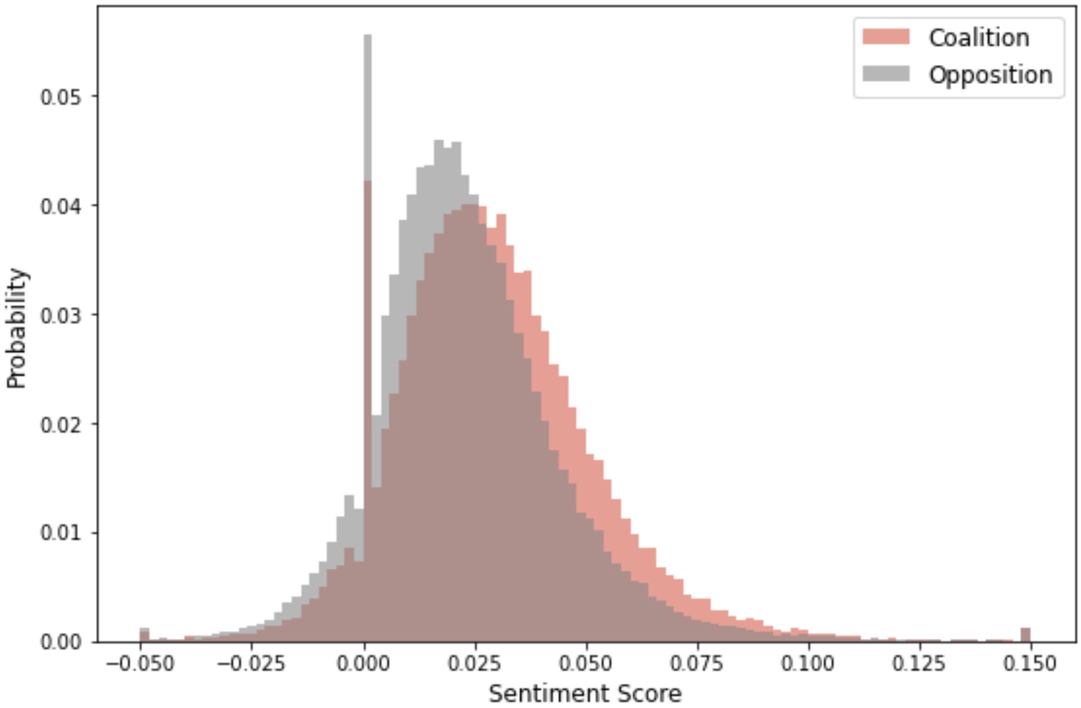

Figure 5.2 Histogram of by-speech scores for coalition and opposition speakers in ParlaMint-DK. The data is clipped at the extremes to fit into the desired window.

Speech scores follow roughly normal distributions around the means, with spikes at 0, as there is a large number of speeches without any sentiment words – the "true neutrals". Because the average speech is positive at 0.025, these neutrals tend to depress the overall score. While coalition and opposition speech scores overlap considerably, there is a higher share of more-positive-than-average coalition speeches, and conversely a higher share of coalition speeches that are less positive than average (including neutral).

II. Sentiment Variance over Time

So far sentiment has been treated as a static measure, but this may be misleading in the presence of diachronic patterns, which have been attested in previous work (Rheault et al., 2016). Fig. 5.3. presents sentiment scores by party status calculated for one month intervals. Overall, sentiment divergence between coalition and others is consistent across the whole time period. During the last government in the data set (Frederiksen I), the "Other" category seems to express more positive sentiment than the opposition, especially during the first part of the term. This indicates the role of "third-party" MPs is not static and may be oppositional or supportive depending on circumstances.

Major outliers appear in June of 2017 and 2018, as well as around the government change in 2019. Closer scrutiny of the agenda titles and speeches during this period reveals that there were very few brief meetings during these months at the edge of term and session changes, resulting in greater variance (sentiment variance over time is naturally also greater on a by-date basis). Autocorrelation scores (Tab. 5.5) are low and there is little evidence of time-dependency, except for the "Other" category – likely due to this group's visible increase in positive sentiment during the Frederiksen I cabinet.

Party Status	Mean	Std	SE	Autocorrelation
Coalition	0,0316	0,0029	0,0003	0,0241
Opposition	0,0241	0,0038	0,0004	0,0955
Other	0,0242	0,0033	0,0003	0,2883

Table 5.5 Descriptive statistics for monthly sentiment scores by party status in ParlaMint-DK.

The COVID-19 period is associated with slightly lower average sentiment for both coalition and opposition speakers (Tab 5.6), indicating a negative emotional response to crisis (Rheault et al., 2016). However, no significant long-term changes in sentiment are apparent, making aggregate measures over time appropriate.

Subcorpus	Coalition	Opposition	Other
COVID-19	0.0313	0.0225	0.0251
Reference	0.0321	0.0238	0.0232

Table 5.6 Sentiment score by party status for the COVID-19 subcorpus (November 2019 – June 2022) and the reference corpus (October 2014 – October 2019) in ParlaMint-DK.

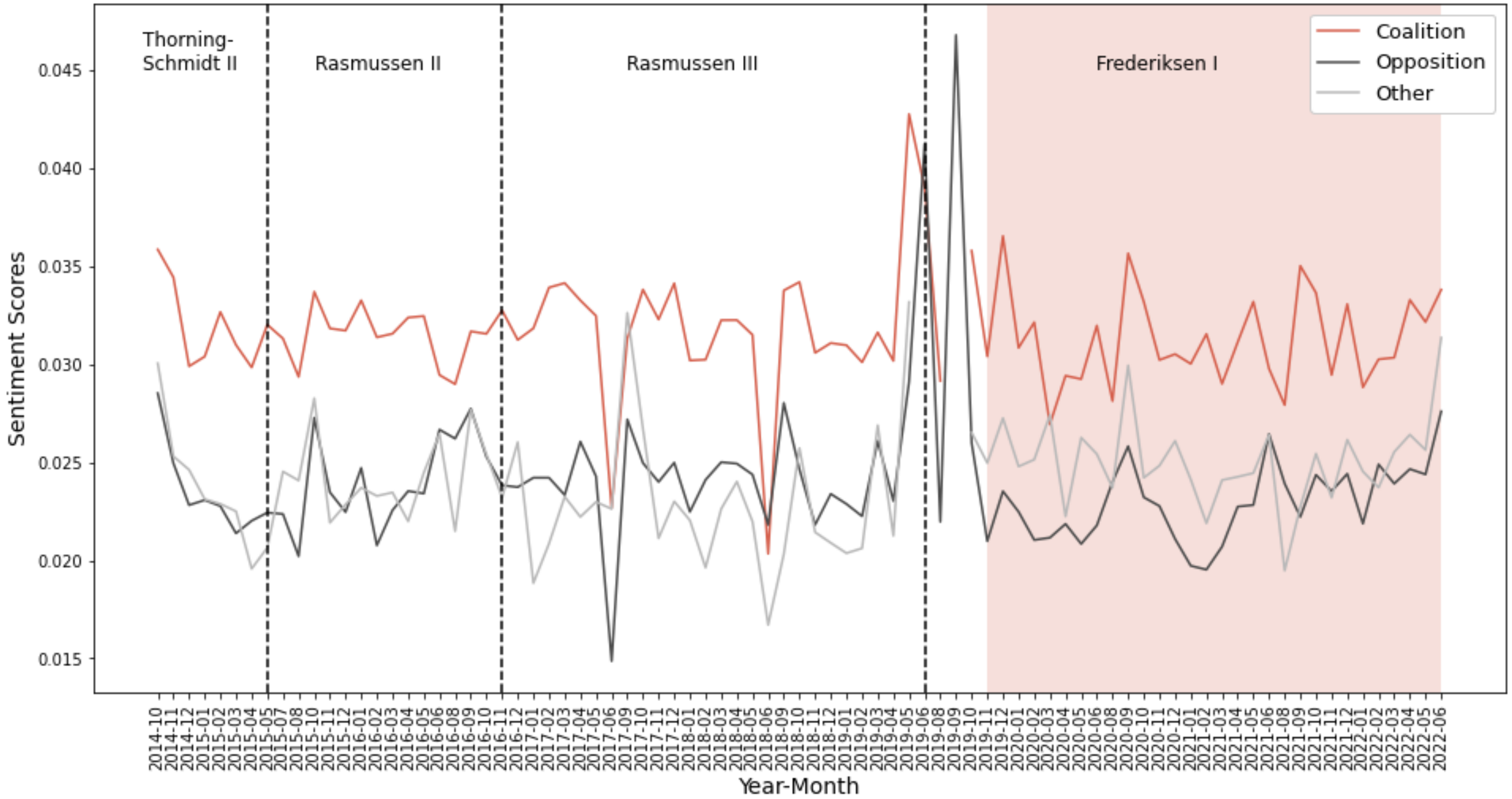

Figure 5.3 Sentiment scores by month and party status for ParlaMint-DK. Not all months of a year are represented, and during some low-activity months, not all party statuses are represented. Dotted lines indicate a change of government, labelled in black. The red background marks the COVID-19 pandemic period.

III. Sentiment and Proceedings

Different proceeding types are also expected to shape emotional polarity for various political actors. Fig. 5.4 shows sentiment scores by party status for the three most common proceeding types in the Danish parliament. Overall, sentiment is highest during negotiations and lowest for questioning, with readings falling in the middle, although differences are relatively small. High negotiation sentiment may reflect the need to appeal to one's opponents and find common ground, though coalition and opposition speaker sentiment is still quite polarized in that context.

The party status sentiment gap is greatest during questioning, likely due to "attack-and-defense" character of this type of speech (see Hypothesis III.). On the other hand, readings have the smallest coalition-opposition gap, which may be indicative of a more collaborative approach to legislating.

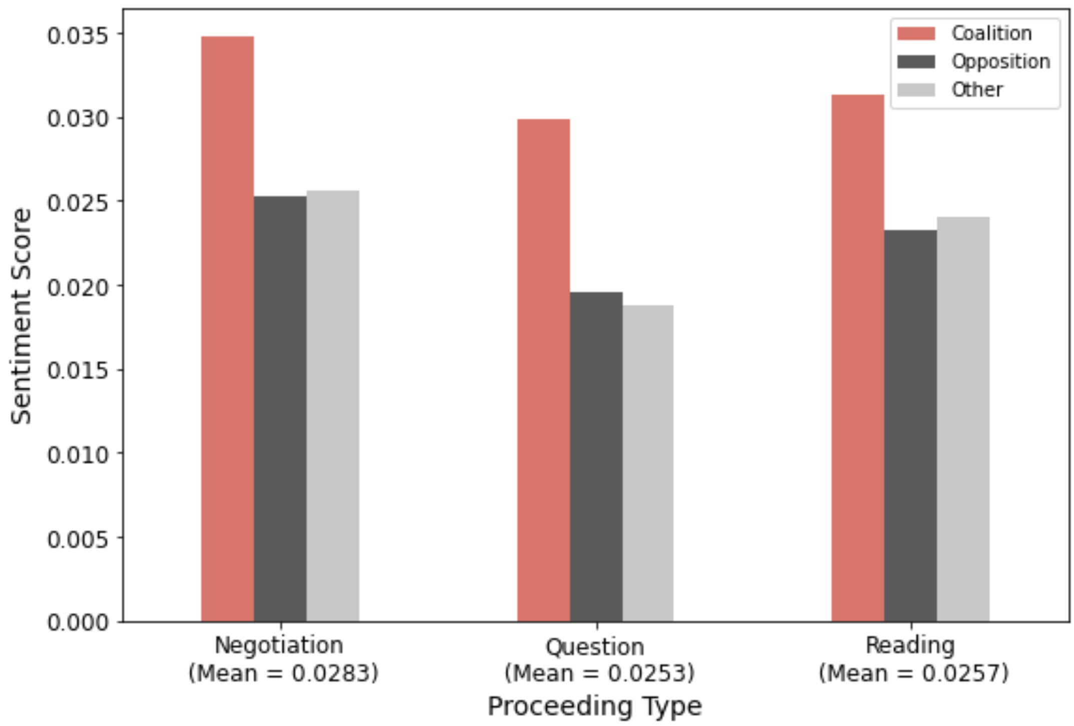

Figure 5.4 Sentiment score by party status and proceeding type in ParlaMint-DK. Averages across party status are marked in parentheses.

5.2.2 The Bulgarian Corpus

I. Pre-Processing

Analyzing the Bulgarian data required some additional pre-processing. For one, caretaker governments were excluded from most of the analysis, as they bias the "Other" and "Opposition" categories.

During a preliminary gold standard annotation, it also became apparent that a non-trivial number of speeches in the Bulgarian parliament consist of the announcement and reading of bill and other proposal texts by members of legislative committees (see Section 3.3.2). These speeches have a purely procedural purpose, and any sentiment words they might contain, e.g., in the title of a draft bill, are accidental and not substantive. Furthermore, it was found that these speeches were overwhelmingly given by coalition MPs, which could potentially skew comparisons by party status.

Therefore, as an extra pre-processing step, speeches marked as "committee" were removed from the corpus before sentiment analysis. These speeches made up made up almost 15% of the total words in the corpus, highlighting their confounding potential.

II. Sentiment and Party Status

Fig. 5.5 illustrates sentiment in relation to party status and gender. Since ministers rarely speak in the Bulgarian parliament, coalition speakers have not been subdivided into regular MPs and ministers. The graph indicates that coalition members express slightly more positive sentiment than opposition speakers, but the differences are very small. Speakers in the "Other" category do not obviously align with either group.

Gender effects are apparent for coalition and opposition speakers, with women expressing more positivity. The difference between coalition and opposition sentiment appears larger for men, with opposition men being the most negative of any group. Interestingly, the "Other" category (independents) shows no gender differences, though conclusions about this group are fraught because of its small size.

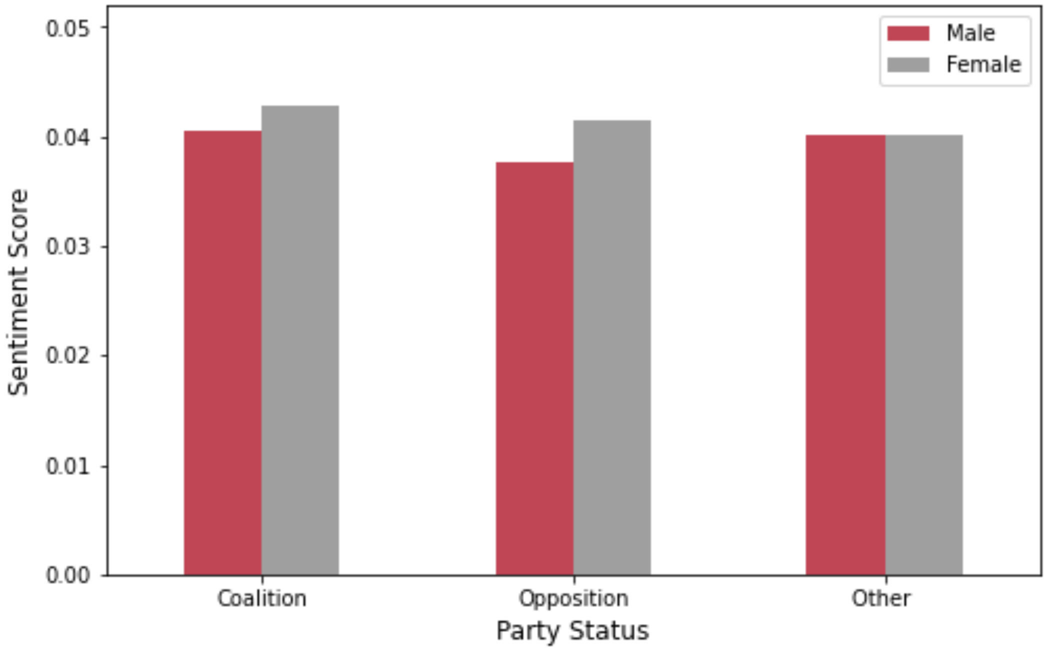

Figure 5.5 Sentiment by party status and gender in ParlaMint-BG.

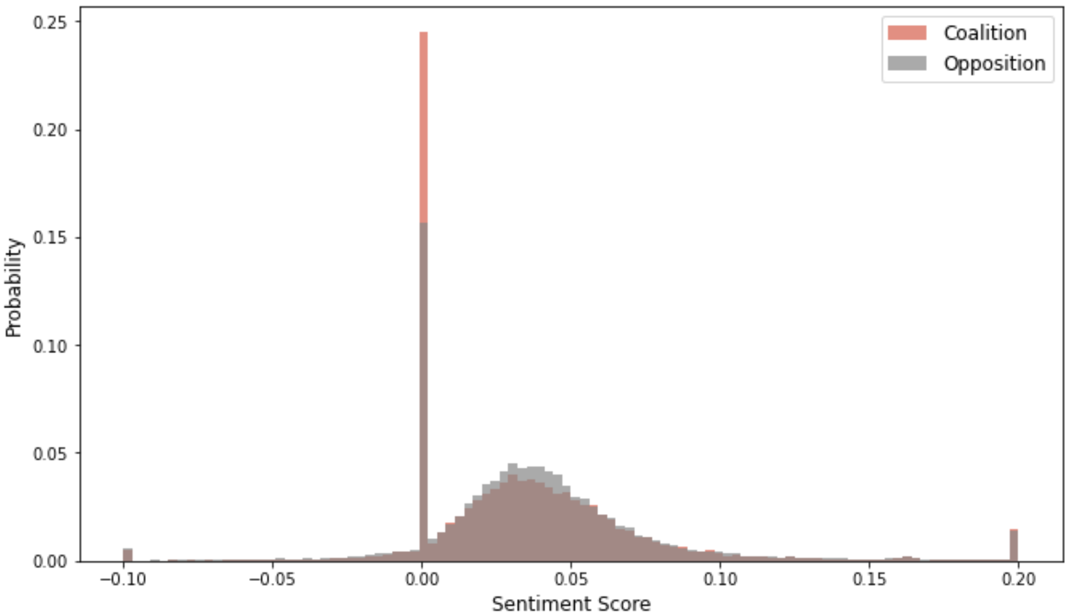

Figure 5.6 Histogram of by-speech scores for coalition and opposition speakers in ParlaMint-BG. The data is clipped at the extremes to fit into the desired window.

A by-speech histogram of sentiment score distributions (Fig. 5.6) shows almost full overlap between coalition and opposition scores, with the opposition distribution being slightly more peaked at the mean. We also see a very large proportion of neutral speeches rated 0 – almost 25% for coalition speakers. With a mean token length of 14,6, these speeches tend to be very short. As only 7% or about 1 in 15 words in the corpus are sentiment words from ParLex, short speeches are expected to be disproportionately neutral.

The histogram results also illustrate why treating individual speeches as documents can be spurious. On a by-speech basis, coalition speakers appear less positive than opposition speakers, because the abundance of short neutral speeches depresses their average speech score (0,035 for coalition and 0,038 for opposition). On a by-word basis, as in the main analysis, it is actually clear that coalition speakers do use more positive expressions in their speech as a whole.

III. Sentiment Variance over Time

When considering variation over time, the by-month time series for sentiment includes caretaker governments (the unmarked periods in between government changes) for completion. The data (Fig. 5.7) appears to exhibit some time-dependency. Speakers in the "Other" category have high sentiment variance, especially during the Borisov III and Petkov governments, but not during Borisov II. This reflects the number of independents in parliament during those respective governments (see Tab. 3.4. in Section 3.2.4), as fewer speakers and words will result in more variant scores. With regard to coalition and opposition speakers, there is no consistent sentiment gap visible at this level of analysis.

Relevant statistics for the by-month scores are shown in Tab. 5.7. Most notably, autocorrelation scores confirm a degree of time-dependency of the data, suggesting that party groupings go through more and less positive periods rather than maintaining a constant average sentiment over time. However, there appears to be no general trend toward higher or lower sentiment as in Rheault et al. (2016).

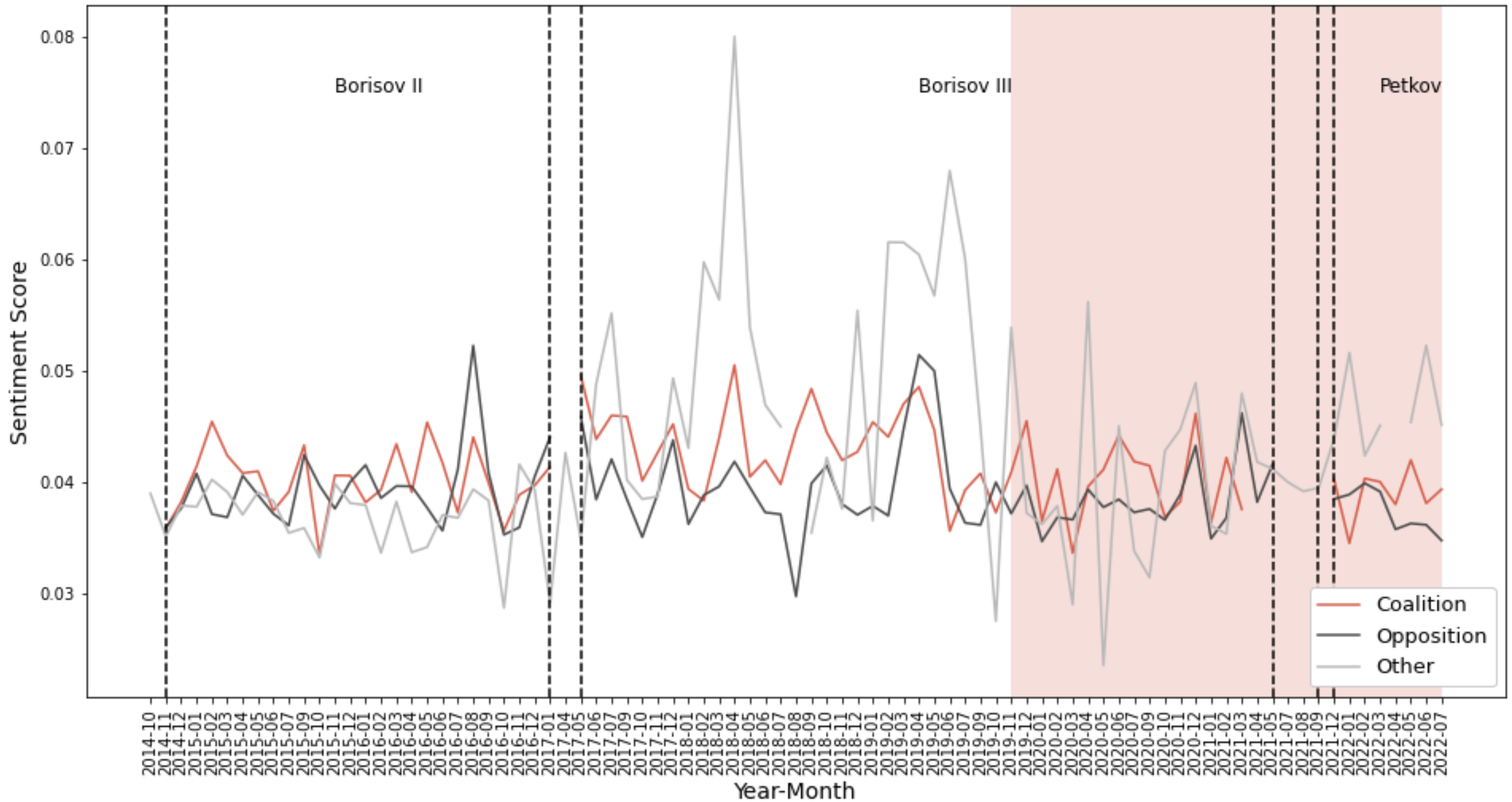

Figure 5.7 Sentiment scores by month and party status for ParlaMint-BG. Not all months of a year are represented, and during some low-activity months, not all party statuses are represented. Dotted lines indicate a change of government, labelled in black. Caretaker governments are not labelled but may be inferred from Tab. 3.4. The red background marks the COVID-19 pandemic period.

Party Status	Mean Score	Std	SE	Autocorrelation
Coalition	0,0411	0,0036	0,0004	0,2017
Opposition	0,0390	0,0036	0,0004	0,2946
Other	0,0426	0,0097	0,0010	0,3942

Table 5.7 Descriptive statistics for monthly sentiment scores by party status in ParlaMint-BG.

Tab. 5.7 also shows that coalition speakers are in fact more positive on average when considering by-monthly scores. The small differences are due to the fact that sentiment divergence by party status is only present for certain types of proceedings (see below) and so will not be apparent during every individual month.

Finally, the COVID-19 subcorpus is associated with lower average sentiment than the reference subcorpus, consistent with the likely increase of negative themes and sentiment expressed in response to the pandemic.

Subcorpus	Coalition	Opposition	Other
COVID-19	0.0397	0.0377	0.0403
Reference	0.0416	0.0386	0.0389

Table 5.8 Average sentiment by party status for the COVID-19 subcorpus (November 2019 – June 2022) and the reference corpus (October 2014 – October 2019) in ParlaMint-BG.

IV. Sentiment and Proceedings

While overall differences in sentiment between coalition and opposition speakers have been very modest in aggregate, Fig. 5.7 demonstrates that sentiment divergence by party status is strongly mediated by effects of proceeding type in the Bulgarian parliament.

The lowest mean score and the smallest differences between coalition and opposition occur within the legislative proceeding type, which encompasses readings and voting (see also Section 3.5.2). This matched sentiment may indicate that the legislative process is fairly cooperative. Due to the bias of the sentiment scale, a low mean score could either be due to more negative or more neutral (e.g., procedural) speech during these proceedings. In contrast, during draft decisions and questioning, there are significant gaps between coalition and opposition speakers. This is consistent with Hypothesis III., as opposition speakers are expected to be critical of government actions, while ministers will maintain a defensive posture. That this pattern also holds for draft decisions confirms the qualitative impression noted in Section 3.2.4 that this proceeding type also tends to be explicitly adversarial.

Coalition speaker sentiment is the most variable across proceeding types, while opposition sentiment appears least variable. This may reflect the fact that navigating disagreements while in power can necessitate a greater emotional and rhetorical "toolbox" than maintaining a consistent oppositional stance.

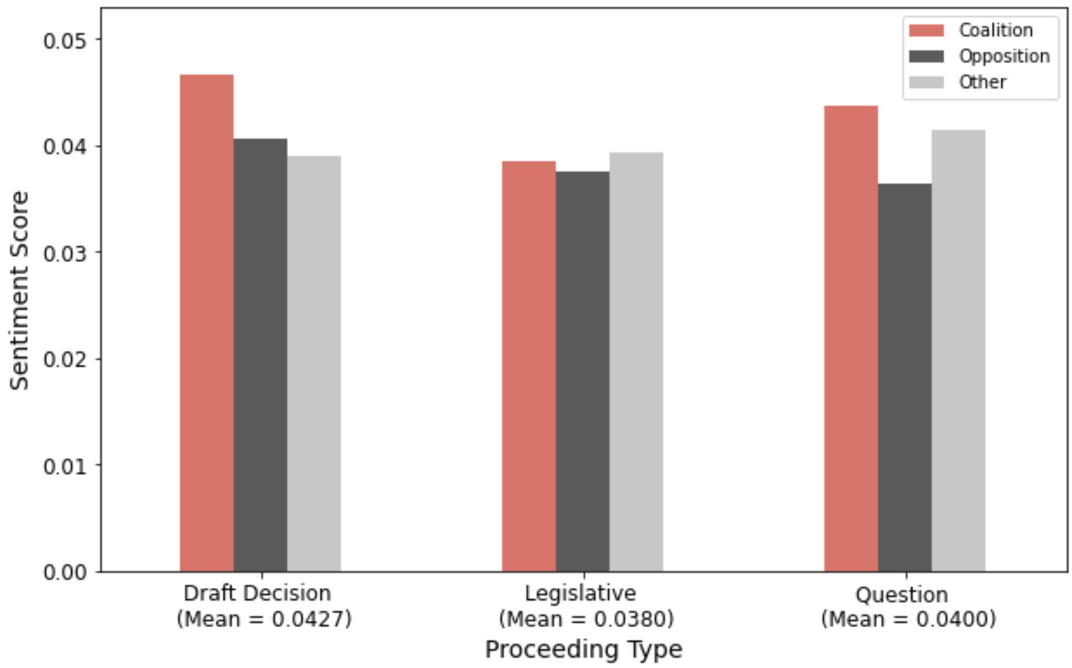

Figure 5.8 Sentiment score by party status and proceeding type in ParlaMint-BG. Averages across party status are marked in parentheses.

5.2.3 Sentiment: Comparative Summary

A sentiment analysis of the two corpora reveals some similarities. In particular, the effect of gender on sentiment is remarkably consistent across the two parliaments and for speakers of different party status: women use more positive language than men. Furthermore, the COVID-19 pandemic period is associated with less positive sentiment in both parliaments.

With regard to the main research questions on the effect of party status and proceeding type, the two corpora diverge. In the Danish parliament, there is a large effect of party status, with coalition speakers measuring as considerably more positive than opposition or other speakers across proceeding types. Speeches during ministerial questioning are the least positive on average.

In the Bulgarian parliament, there is only a very small overall effect of party status in favor of coalition speakers. However, when broken down by proceeding type, we see a more pronounced coalition-opposition divergence during discussions of draft decisions and questioning, while the gap remains small during legislative meetings, which also tend to be the least positive overall.

6 DISCUSSION

This chapter evaluates and critically discusses the results of the study. Section 6.1 presents a quantitative evaluation of results, including inter-annotator reliability for sentiment (6.1.1), the correspondence between automatic methods and human judgment (6.1.2), and a performance comparison for different lexica (6.1.3). In Section 6.2, results are considered in light of the research questions and hypotheses in interaction with previous research. Finally, Section 6.3 discusses methodological limitations and alternatives.

6.1 Evaluation

As NLP methods for analyzing political speech have grown in popularity, researchers have stressed the importance of validation, particularly with unsupervised methods such as scaling or topic modelling (Chen et al., 2023; Grimmer & Stewart, 2013). Since the sentiment analysis performed for this thesis also hinges on an unsupervised technique for finding domain-specific sentiment words, it is important to evaluate the reliability of the results before drawing final conclusions. Following studies such as Proksch et al. (2019), sentiment results will be validated based on a hand-annotated gold standard, comprising 300 annotated speech excerpts for each parliament. Unlike other common validation methods, like movie review classification (as in Rheault et al., 2016), this approach will allow for an assessment of the advantage of domain-specificity, the distinguishing feature of ParLex.

In this section, construct validity will therefore be investigated by estimating inter-coder reliability for sentiment as well as the correlation between automated and hand-annotated scores. ParLex will then be compared to other available sentiment lexica to judge whether domain-specificity improves performance for parliamentary speech. This section especially pertains to Research Question and Hypothesis I.

6.1.1 Inter-Annotator Reliability

When developing automatic methods for tasks such as sentiment analysis, human judgment is considered the upper bound for performance. However, even commonplace concepts can have fuzzy perceptual correlates. To investigate how well sentiment is defined for parliamentary speech, two annotators were recruited to hand-code 100 of the speeches in the gold standard for comparison with the annotations produced by the author.

In cases of disagreement, it is common practice for annotators to discuss and agree on a score, but this type of back-and-forth was not feasible. Instead, disagreements were reviewed by the author and the original ratings were amended in cases where the other coder was judged to have been more accurate at second glance. This means that inter-coder reliability may be underestimated, as the secondary coders did not get the opportunity to review their own ratings.

Inter-annotator reliability is estimated by the Cohen Kappa score (Cohen, 1960) as implemented in Scikit-learn (Pedregosa et al., 2011). Unlike simple agreement measures, Cohen's Kappa accounts for the probability of chance agreement based on the number of categories. The Kappa is first computed for the full five-point sentiment scale (see section 3.4), and then for a three-point scale, in which the two positive and two negative categories, only distinguished by intensity, are combined. Tab. 6.1 shows the results for both corpora, and Fig. 6.1 presents the confusion matrices for the three-way contrast.

Parliament	κ_3	κ_5
Bulgaria	0,78 (substantial)	0,71 (substantial)
Denmark	0,66 (substantial)	0,40 (fair)

Table 6.1 Cohen's Kappa for sentiment annotation on the Bulgarian and Danish parliaments. Kappa subscripts indicate the number of categories used. Labels interpreting the degree of agreement are taken from Viera & Garrett (2005).

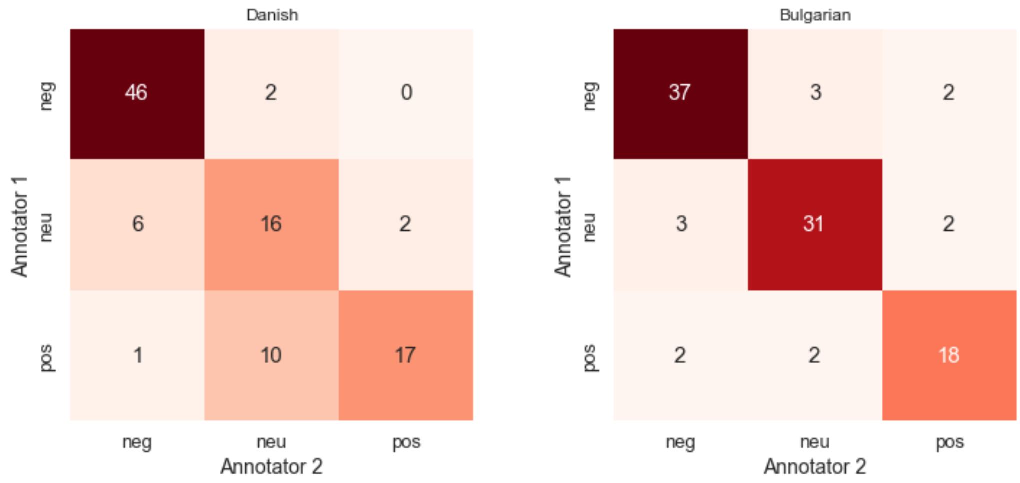

Figure 6.1 Confusion matrices for annotator agreement for Danish (left) and Bulgarian (right) speeches. Annotator 1 is the author in both cases.

Bulgarian agreement is consistently "substantial", while the Danish κ_5 score is significantly lower than the κ_3, indicating uncertainty about sentiment intensity. Even with a three-way contrast however, there is less agreement about the Danish speeches. Fig. 6.1 indicates that this is due in particular to a larger number of annotator disagreements for neutral and positive ratings. The relevant speeches, reviewed manually, tend to be statements of intention and approval with positive framings. The secondary annotator may have considered the effect of those framings rather weak, as a lot of the content in the speeches is purely factual. In general, members of parliament may refrain from very explicit and exuberant praise of their own proposals, preferring to justify them through factual reasoning in order to appear less partisan, giving rise to this ambiguity. Annotator disagreements may also be colored by the fact that the author speaks Danish as a second language, resulting in somewhat different intuitions about interpretation.

On the other hand, there is little disagreement for negative ratings in either corpus. This may indicate that negativity is more perceptually salient and well-defined than positivity, at least for this genre of speech (see also Young & Soroka, 2012). The proportion of speeches rated negative by agreement is also the largest, in contrast to the average positive score that the automated methods find.

Overall, the agreement scores indicate that, while some uncertainty exists in human judgment, sentiment is fairly well-defined for parliamentary speech.

6.1.2 Validation

A major question for sentiment lexica, and affective computing in general, is whether the results they produce agree with human perception. While it has already been established that hand-annotated score distributions tend toward the negative, while automated ones are positive on average, the quantity of interest in this study has not been absolute, but relative differences in sentiment. For this purpose, it is enough to establish a strong correlation between automated and human scores.

When comparing continuous (automated) and ordinal (hand-coded) data, it is better to have larger ordinal scales to approximate continuousness, so the five-point scale will be used. Fig. 6.2 shows the results of the correlational analysis. We see a clear trendline with moderate correlation, but there is also a lot of variance at each level. This indicates that, while any individual speech score is likely to be quite uncertain, large sample sizes, such as the ones used in the sentiment analysis, should yield meaningful differences in average sentiment. When testing at the sample size of the gold standard (Tab. 6.2), party status differences were directionally similar, while gender differences were directionally reversed. However, only one contrasting pair reached statistical significance[16] in a t-test analysis, indicating the necessity of more data for full validation.

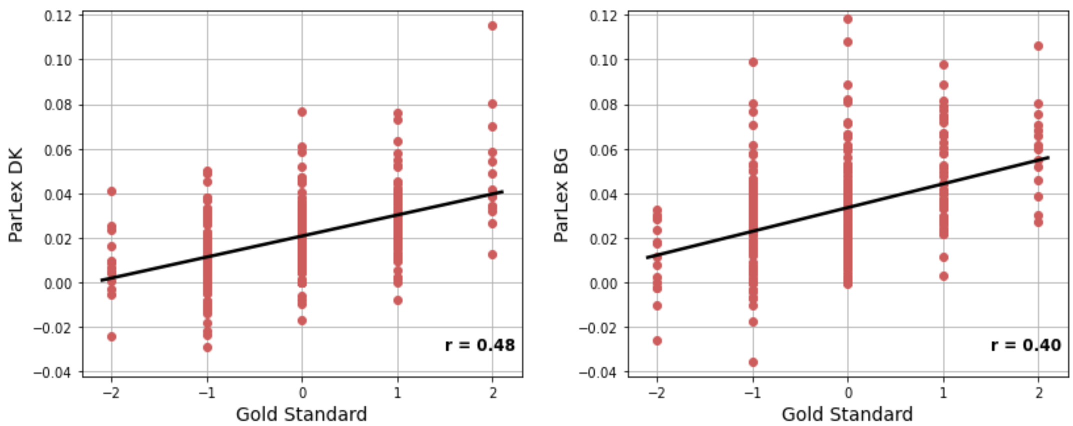

Figure 6.2 Correlation between ParLex and hand-annotated scores for Danish (left) and Bulgarian (right). Correlation coefficients are listed in the bottom right.

[16] Because statistical significance is only meaningful on a by-speech level of analysis, it has not been reported elsewhere. An exploratory t-test analysis revealed that practically every investigated difference in ParLex sentiment is highly significant for the full sample size (50-100K speeches).

Parliament	Annotation	Opposition	Coalition	Male	Female
Bulgaria	Auto	0,0306	0,0342	0,0316	0,0317
	Hand	-0,2590*	-0,0175*	-0,1363	-0,3893
Denmark	Auto	0,0185	0,0231	0,0191	0,0219
	Hand	-0,1583	0,0439	-0,0845	-0,1059

Table 6.2 Average hand annotated and ParLex sentiment scores across party status and gender. The more positive sentiment within each contrasting pair is shaded. Significance is indicated with an asterisk.

In conclusion, there is a moderately strong correlation between hand-coded and automatic scores, allowing a reasonable degree of confidence in the study results. A larger gold standard would be needed to reproduce specific effects.

6.1.3 Domain-Specificity

Using word embeddings to create ParLex was intended to capture the specifics of sentiment within the rather unusual genre of parliamentary speech, following Rheault et al. (2016). However, the study authors never test whether this technique actually outperforms general-use lexica. To determine the advantage of domain-specificity, the Danish version of ParLex was tested against AFINN (Nielsen, 2018) on the gold standard. For Bulgarian, no quality language-specific sentiment lexica exist. A publicly available sentiment dataset developed by Chen & Skiena (2014), generated via graph propagation and machine translation based on English sentiment lexica, is available in 81 languages, including Bulgarian, and was therefore used for this purpose[17].

Implementing the two lexica was simple. The gold standard speeches were tokenized and lowercased. AFINN contains most word forms for each lemma, so no lemmatization was necessary. The Bulgarian lexicon is not morphologically complete, but its authors indicate that it is intended for direct analysis on tokenized text (Chen & Skiena, 2014). Negations are accounted for in the same way as for the ParLex analysis, by neutralizing words in between negation words

[17] The original intention was to translate the Lexicoder Sentiment Dictionary (Young & Soroka, 2012), but a scarcity of time and lexical resources for dealing with the stemmed format of lemmas in the LSD prevented this.

and punctuation – and this step does measurably improve correlation with hand-coded scores.

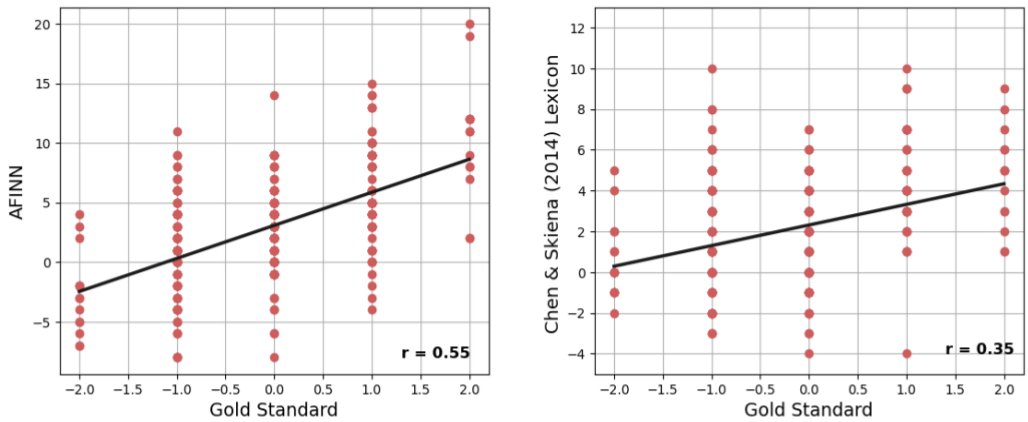

Figure 6.3 Correlation between hand-annotated and general-use sentiment scores for Danish (left) and Bulgarian (right). Correlation coefficients are listed in the corner of each plot.

Fig. 6.3 presents the correlation between the general-use lexica and the gold standard annotations. The Danish AFINN lexicon significantly outperforms ParLex in this regard, improving the correlation coefficient by 7 points. The Bulgarian lexicon from (Chen & Skiena, 2014) fares worse than ParLex BG by 5 points. However, even this result is impressive considering the morphologically incomplete nature of this lexicon.

The overlap between ParLex BG and the Chen & Skiena (2014) lexicon is 555 lemmas, comprising 39% of words in ParLex. The correlation between the sentiment scores they produce is low (r = 0.18). For ParLex and AFINN, the overlap is 469 words, which amounts to 33,5% of ParLex. The scores of these two lexicons, on the other hand, have a more substantial correlation of r = 0.5.

Overall, these results indicate that domain-specificity does not present a clear advantage. Where high-quality general-purpose lexica exist, they can validly be used on parliamentary speech. Expanding a seed list of sentiment words using embeddings can however be a beneficial solution in the absence of sentiment resources, as for Bulgarian. Furthermore, based on the low degree of overlap between lexica, an expansion strategy could potentially serve as a way of extending existing resources rather than creating new ones from scratch. Alternatively, a “mixture of experts” approach, whether rule-based or machine learning, could combine and weigh sentiment scores produced by different lexica to determine the final result.

6.2 Interpretation

In this section, study findings are summarized, interpreted, and discussed in light of previous research. Research questions II.-IV. and their associated hypotheses are examined in particular.

6.2.1 The Meaning of Sentiment Divergence

Despite the fact that coalition and opposition MPs may relate very differently to each other in different parliaments (De Giorgi & Marangoni, 2015), work from Proksch et al. (2019) indicates that sentiment can reliably distinguish speakers by party status for all seven European countries in their sample. Consequently, Hypothesis II. anticipated a positive sentiment difference in favor of coalition speakers, as they are likely to support legislative proposals and speak defensively during hearings.

Broadly speaking, the results of this study confirm the hypothesis and accord with previous work. Across both Bulgarian and Danish parliaments, coalition speakers are more likely to use positive language. However, the effect is much more consistent in the Danish Folketing, where the divergence persists across all different proceeding types and is of a significant magnitude. In Bulgaria, the average effect of party status is small and strongly mediated by proceeding type, being most pronounced during ministerial hearings and draft decisions. During legislative readings and debates, coalition and opposition speakers appear quite matched in emotional tone.

Sentiment-wise, speakers in Danish non-coalition "supporting parties" actually tend to align with opposition speakers across all proceeding types, indicating that their role is perhaps more complex than the name indicates. In the Bulgarian parliament, independents are few and speak rarely, so their results are subject to a lot of uncertainty. They do not clearly align with either coalition or opposition speakers on sentiment.

These results complicate existing interpretations of sentiment divergence as presented in Proksch et al. (2019), who see it as an expression of opposition critique towards government-proposed bills and therefore an essential measure of legislative conflict. The challenge is two-fold. Firstly, this effect is not clearly apparent in the Bulgarian parliament during legislative meetings, despite the generally adversarial character of Bulgarian political speech (Tarasheva, 2014). Second, sentiment divergence by party status is in fact more pronounced during non-legislative meetings such as questioning, negotiations and procedural draft decisions – contexts that are less dominated by government proposals than legislation. In general, treating legislative debates as the primary site of interest may obscure the more multi-faceted expression of political conflict in different parliaments.

Instead, while legislative critique plays a part in sentiment divergence, it is also an expression of supra-ideological political conflict, called language of "attack and defense" in Hirst et al. (2010). This explains why coalition-opposition gaps (Tab. 6.3) in both parliaments are most pronounced during ministerial hearings, when coalition speakers are most likely to have to defend themselves against explicit accusations and attacks. Gaps are smallest during bill readings and legislative debate, which may require somewhat more persuasion and subtlety in order for the various parties to achieve their goals.

Parliament	Legislative	Negotiation	Draft Decision	Questioning
Bulgaria	0,001	-	0,005	0,007
Denmark	0,008	0,009	-	0,010

Table 6.3 Average sentiment difference between coalition and opposition speakers across proceeding types for both parliaments. The difference is in favor of the coalition.

There is also the question of why Bulgarian sentiment gaps, particularly during legislative debates, are smaller than Danish ones, contrary to Hypothesis IV.a. While Danish party positions and ideologies are well-defined, Bulgarian parties are less ideologically consistent and often headed by charismatic individuals with populist leanings (Karasimeonov, 2019). The ideological well-formedness and party discipline of Danish politicians (Green-Pedersen & Skjæveland, 2020) may cause them to act and speak more consistently with their roles as coalition and opposition. Because of their eclectic ideologies and high turnover rate (Kolarova, 2018), Bulgarian MPs may conversely have less regular partisan speech patterns on legislative policy matters. The many political and personal conflicts between various Bulgarian speakers appear to be hashed out during hearings and draft decisions instead.

One last observation bears notice with regard to proceeding types. In Hypothesis III., it was speculated that, because of their adversarial character, questionings/hearings would prove to have the lowest overall sentiment scores. This holds in the Danish parliament but is only true for opposition speakers in Bulgaria. Because coalition speakers will "compensate" by defending themselves in such contexts, the overall average does not decrease much. In general, it is not meaningful to compare proceeding types without also breaking down speakers by party status, as coalition and opposition speakers may have very different incentives and speech patterns in different contexts.

6.2.2 Gender Differences

During the sentiment analysis, scores were broken down by gender as well as party status in order to account for a possible confounder, even though gender was not a research focus. A consistent sentiment gap in favor of women was found across both parliaments and all party statuses (except independents in the Bulgarian National Assembly), though the effect is relatively small.

So why do women appear to express more positive sentiment? The pattern is consistent with prior research on emotional polarity and gender (Haselmayer et al., 2022). One interpretation ascribes this dynamic to a role congruency effect, in which women's expression is responding to the societal expectation that women be more agreeable and communal-minded, including in parliamentary contexts (Banducci et al., 2012; Hargrave & Langengen, 2021). When women act incongruently to assigned roles by using negative language, it can have punitive consequences (Brescoll & Uhlmann, 2008). Although recent research indicates that rising shares of female MPs and changing social mores may be closing the gender gap in expressions of e.g., aggressiveness (Hargrave & Blumenau, 2022), it is striking that this pattern holds across two countries with very different gender politics and attitudes.

On the other hand, the difference is small and may also be due to some particularity in the construction of ParLex. Women and men have different speech patterns apart from sentiment (Dahllöf, 2012; Mandravickaitė & Oakes, 2016), e.g., in the use of function words, some of which are present in ParLex and possibly acting as confounders. Closer investigation of the sentiment keywords driving the divergence between men and women could potentially illuminate this problem, although it is beyond the scope of this study.

6.2.3 Cross-National Comparison

The goal of the format standardization of the ParlaMint project is facilitating cross-national and cross-lingual comparisons. A major study from the project, Miok et al. (2022), presents a methodology for analyzing ParlaMint corpora with a variety of NLP techniques. They caution, however, that "[interpreting] the results require[s] interdisciplinary collaboration and understanding of language and political situation in the countries tackled" (p. 23-24, Miok et al., 2022). Implementing the technique and interpreting results from ParLex are both language-dependent and difficult to fully automate. Another challenge, noted by Proksch et al. (2019), is anchoring measures of sentiment objectively so that direct national comparisons can be made. They attempt to do this by using the same translated sentiment dictionary for all parliaments.

The ParLex methodology results in even less objective grounding, as both sentiment dictionaries are essentially "normed" by their respective corpora. Making the seed dictionaries as close to identical as possible could have mitigated this fact. Instead, at two points in the lexicon creation process, lemmas were filtered by frequency, resulting in different biases and the inclusion of a number of frequent "junk" words, particularly in positive sentiment lists. Relying less on frequency may have improved the quality of the dictionaries in identifying sparser, but more accurate sentiment words.

All of this means ParLex is not truly cross-lingual, which makes cross-national comparisons difficult. Hypothesis IV.c, for example, anticipated more emotional language in the Bulgarian parliament than the Danish one. While there are in fact slightly more sentiment words in the Bulgarian corpus, it is difficult to say whether this difference is meaningful or merely an artifact of frequency cutoffs in ParLex. Vries (2022), who also broadly follows the methodology of Rheault et al., (2016), introduces an intermediate validation step using hand-annotated sentences to find the ideal cutoff points for inclusion in the final sentiment dictionary. This type of hyperparameter tuning requires additional manual annotation, but it could have ensured better psychological validity of the sentiment measure.

Hypothesis IV.b also predicted more negativity on the part of Bulgarian parliamentarians. In fact, both parliaments had 85-15 splits in favor of positive sentiment words based on the ParLex analysis, a surprising similarity that does not accord with prior knowledge about the different styles of politics in the two countries. Manual annotation scores from the gold standards, on the other hand, do rate Bulgarian speeches almost twice as negatively as Danish ones on average (see Tab. 6.2). This limitation on the part of ParLex may be due to limited sensitivity to sentiment intensity, which is only modelled as vector similarity to the original seed words. The qualitative keyword analysis did not indicate that Bulgarian negative terms were more intensely negative. However, intensity is contextual and often expressed in more complex syntactic constructions that a mere lexical analysis cannot capture. Adding syntactic rules or otherwise improving intensity weights is another avenue for the improvement of ParLex.

Despite these difficulties, there are still interesting cross-national findings to present. The qualitative keyword analysis revealed many common themes in positive and negative speech, indicating a fairly consistent rhetorical style among MPs of both nationalities. The topical themes found in keywords, on the other hand, revealed some differences in political priorities between the two countries. For example, two distinct themes relating to immigration and multiculturalism were found in the negative Danish keywords – consistent with the prominence and controversy of this topic in Denmark in recent years (Navarretta et al., 2022). On the other hand, the negative Bulgarian list included a topical theme related to corruption and embezzlement scandals. While these results are interesting and consistent with prior knowledge about both countries, a joint sentiment and topic modelling analysis (in the vein of Abercrombie, 2021) could probably elucidate many more details.

The extraction and analysis of proceeding type labels also revealed striking differences in the amount of speech and time allocated to different parliamentary tasks. The Danish parliament is characterized by substantive debate during bill readings as well as frequent policy negotiations. By contrast, there is much less legislative and policy debate in the Bulgarian parliament, where “draft decisions” concerning committee appointments or changes to the parliamentary rulebook constitute a large part of parliamentary speech. During the manual annotation process, it was also apparent that procedural comments, challenges, and questions abound in the Bulgarian parliament, indicating frequent renegotiation and breaching of norms. These “procedural” speeches in the gold standard were labelled and made up 10,6% of the total, while no speeches were labelled as procedural in the Danish gold standard. The procedural focus in the Bulgarian parliament matches the contrast found in relation to the British House of

Commons by Tarasheva (2004). A strong emphasis on legislative debates, as in parts of Proksch et al. (2019), may therefore be less appropriate for the study of some parliaments.

6.3 Methodological Limitations

Sub-optimal evaluative results and difficulties in interpretation are likely connected to the methodological limitations of ParLex, partially touched upon earlier. Some of these are inherent to lexicon methods in general. While the valence of particular words can index sentiment, meaning is also expressed syntactically and pragmatically. This is especially true of parliamentary speech, which is formal and syntactically complex, suggesting that a lot of emotional language will go unaccounted-for with lexical tools. The only syntactic feature ParLex takes into consideration is negation, and that relatively crudely. Some lexicon implementations, such as the LSD (Young & Soroka, 2012) and VADER (Hutto & Gilbert, 2014) do include more detailed syntactic rules, though these have only been developed for English, and applying something similar for Danish and Bulgarian was not feasible within the time constraints.

ParLex is also largely automatically generated via embedding similarity, so not all lemmas are hand-validated. There appear to be many words in the sentiment lists that are not necessarily emotionally charged in themselves but are a consequence of the statistical logic of co-occurrence methods like GloVe. For example, "tæt" (*close*) is a relatively frequent "positive" word in the Danish corpus, likely because of its association with the very frequent "samarbejde" (*collaboration*). On the one hand, "close collaboration" does indeed seem more positive than simply "collaboration", though it will not have a clear valence in other contexts, e.g., "we are close to the end of the term", and introduce error. On the other hand, if it does so often intensify "collaboration", a close statistical relationship may in some cases outweigh the generated error – but in others, it will not. In the case of ParlaMint-DK, a hand-labelled sentiment lexicon like AFINN (Nielsen, 2018) outperforms the statistical ParLex, though correlation with hand-coded ratings is merely moderate in both cases.

An alternative to lexicons is methods based on neural networks that can learn more complex relationships between words with labelled data. However, not many labelled datasets exist for parliamentary speech (though see Abercrombie & Batista-Navarro, 2018, for the British House of Commons and Mochtak et al., 2022, for ex-Yugoslavian legislatures). Miok et al. (2022) get around this by training a multilingual BERT model (Devlin et al., 2018) on a diverse set of sentiment-labelled data in different languages, enabling the model to learn representations in a language-independent semantic vector space. The accuracy scores for sentiment classification are fairly high at 85% for Bulgarian (Miok et al., 2022), but only with very conservative cutoffs. Ultimately, both ParLex and this approach try to compensate for the lack of high-quality resources with modest results. The creation of more sentiment-annotated parliamentary datasets may be necessary for real improvement in this area.

6.3.1 Embedding Quality

The performance of ParLex also greatly depends on word embedding quality during the lexicon expansion phase. As a general observation, the Bulgarian embeddings seemed less successful than the Danish ones in capturing relevant lemmas. For example, the large number of "positive" function words in the Bulgarian expansion list motivated the final inclusion of a frequency criterion, even though this problem was not conspicuous in the Danish lists.

Without quantitative measures, however, it is difficult so speak about this precisely. Measuring embedding quality would also permit a degree of hyperparameter tuning for vector sized or minimum vocabulary frequency. One way to accomplish this would be the validation step from Vries (2022) described earlier. Alternatively, Stanford NLP's GloVe algorithm provides a fairly extensive evaluation module consisting of word analogy tests for topics such as family, countries, and grammatical categories, though it is only available in English. Adapting such tests for other languages, and potentially for specifics of parliamentary speech, could prove useful for improving embeddings. The good results achieved by Rheault et al. (2016) with their GloVe embeddings also indicate that the size of the training data matters greatly – with about a century of British Hansard transcriptions, they can draw on vastly more statistical relationships between words. This may also explain the under-performance of the Bulgarian embeddings relative to the Danish ones, as the ParlaMint-BG corpus contains about 2 million fewer words than ParlaMint-DK.

Another major limitation of GloVe is its insensitivity to context and polysemy. This was only partially overcome by adding part-of-speech labels to lemmas during the training process and addressing it fully would require neural embeddings. Transformer-based language models like BERT are the most popular choice when working with sentiment and/or multilingual data (Miok et al., 2022;

Rudkowsky et al., 2018; Yang et al., 2019). Tang et al. (2016) create so-called "sentiment embeddings" by training with sentiment-labeled data to ensure greater similarity between emotionally congruent words. While these types of models require a lot of training data, fine-tuning pre-trained models on parliamentary speech could still provide some measure of domain-specificity without sacrificing general ability. However, this process would still be computationally expensive and necessitate on-line use during inference. BERT models also have input limits that would prevent longer parliamentary speeches from being analyzed in their entirety. For all its limitations, the ParLex approach is lightweight and easy to implement once sentiment dictionaries have been created. Nevertheless, exploring neural solutions is a natural next step for parliamentary sentiment analysis.

6.3.2 ParlaMint Encoding

Comparing parliaments cross-nationally can be challenging because of differently defined roles and institutional customs, but also because of varying transcription quality. While ParlaMint aims to standardize as much as possible, the source data still affects research possibilities. For example, the lack of section headings in ParlaMint-BG was a major setback in this project, and the workaround using chairperson announcements is not completely accurate and does not cover the whole corpus. Furthermore, encoding errors (e.g., the speaker role of Kiril Petkov) and inconsistent scripts (Cyrillic vs. Latin) for some features in ParlaMint-BG posed practical challenges.

Taking full advantage of all the metadata in taxonomy files was also difficult. While existing ParlaMint scripts for extracting metadata are very useful, information such as speaker role categories could have benefitted from greater specificity (such as party leaders, committee members etc.). Differences in the format and placement of speech and word ID's also made coding custom-made scripts laborious at times. Fully standardizing formatting, generating agenda titles, and potentially providing relevant background information for each parliament could improve the usability and accessibility of ParlaMint to an even greater extent. Researchers interested in working with this data should also be aware of these limitations and spend time looking through the data to familiarize themselves with it as a necessary preliminary.

7 CONCLUSION

7.1 Summary of Work

The overall objective of this research was the comparative analysis of the Bulgarian and Danish parliamentary transcriptions with regard to sentiment in different political roles and contexts (Hypothesis I.). This goal has been achieved by the use of the ParlaMint datasets and the creation of domain-specific sentiment lexica (ParLex BG/DK) for estimating sentiment from text, as well as the annotation of a gold standard for validating and evaluating the analysis. An introductory overview of both corpora was also presented in Section 3.4 and may be of interest to future researchers working with ParlaMint-BG and ParlaMint-DK.

For the development of ParLex, initial seed lists of positive and negative lemmas for both languages were manually created through recursive search in WordNets, taking advantaged of detailed networks of synonymity. The ParlaMint corpora were then embedded using the GloVe algorithm, and the initial seed lists were extended with lemmas from the corpora using measures of vector similarity. With the final dictionaries in place, the two corpora were tagged for sentiment on a by-word basis. Sentiment was estimated in a bag-of-words fashion and normalized by document length, accounting for negation. An analysis of the most common sentiment keywords in each corpus revealed cross-national positive themes of security, partnership, and aspiration, as well as negative themes of danger, conflict, harm, and loss.

To aid in analysis, existing speech metadata was extended by extracting features such as agenda titles, proceeding types, ruling governments, and left-right blocs. A novel method for generating Bulgarian agenda titles from chairperson speeches was presented in the process.

The effects of party status, gender and proceeding type on measured sentiment were then explored. As predicted in Hypothesis II., government and coalition speakers were consistently more positive than opposition speakers, across both parliaments. A negative effect of hearings/questioning on sentiment held for all Danish speakers but was only true for opposition speakers in Bulgaria, implying a partial confirmation of Hypothesis III.

Hypothesis IV.a anticipated a greater and more consistent coalition-opposition sentiment divergence in Bulgaria rather than Denmark. In contrast, Danish sentiment gaps were found to be very pronounced across all proceeding types, while Bulgarian gaps were smaller and mostly evident during hearings and draft decision debates and negligible during regular legislative debates. This was interpreted as a reflection of Danish party discipline and more established political norms. On the other hand, the turbulent Bulgarian political environment was expected to result in more emotional (Hypothesis IV.c) and particularly negative (Hypothesis IV.b) speech. While Bulgarian MPs do use slightly more sentiment words as measured by ParLex, the analysis did not indicate that they express more negativity than their Danish counterparts. However, greater Bulgarian negativity was clearly evident in manual annotation scores.

In general, the psychological validity of ParLex sentiment was found to be limited. Its correlation with human-annotated scores was only moderate, despite the relatively high degree of inter-annotator agreement of the gold standard. ParLex also has a strong positive bias which does not accord with the negative bias of human annotators.

Overall, some caution is advised when applying lemma-expansion to generate domain-specific sentiment dictionaries. Embedding quality is important but difficult to monitor. Furthermore, comparisons with existing general-use lexica indicate that the advantages of domain-specificity are limited. Lexicon expansion may be a decent strategy in the absence of other quality lexica, but it does not improve upon fully manually validated sentiment dictionaries.

7.2 Future Application

Despite limitations, this work presents some stimulating implications and directions for future research on legislatures and NLP.

7.2.1 Parliamentary Priorities

A major insight of the comparative analysis was the difference in parliamentary priorities reflected in the types of proceeding that speech is allocated to. Previous studies (e.g., Miok et al., 2022; Proksch et al., 2019) do not account for proceeding type when comparing parliaments cross-nationally, even though the proportion of legislative to other proceedings can vary widely and significantly influence the relationship between coalition and opposition sentiment. Future studies should always consider this a potential confounding variable. The proportion of legislative to procedural speech may also be of direct interest as a potential index of parliamentary function or political stability in different countries, illuminating politico-cultural differences.

7.2.2 Controversial Topics

The keyword analysis in Section 5.1 revealed that some political topics like immigration or corruption are closely associated with negative speech. Topic-specific sentiment analysis applying techniques such as LDA (Blei et al., 2003) or BERTopic (Grootendorst, 2022) to parliamentary speech could further elucidate the relationship between emotional language and political controversy in different countries. In particular, investigating which topics are associated with the most divergence in sentiment across parties or blocs could pinpoint the drivers of political polarization (Fiva et al., 2021; Hansen et al., 2018; Monroe et al., 2017), a rising concern in many developed nations.

7.2.3 Accessibility

The ParlaMint project has a gone a long way in making parliamentary speech data more accessible to researchers. However, it is still difficult for regular citizens to wade through this data. One avenue for future development would be the creation of digital platforms for users to explore e.g., sentiment and keyword breakdowns for parliamentary meetings by date or category. Visualizations could summarize information and provide easy overviews of data for research or informational purposes, helping keep citizens informed and governments transparent. Gkoumas et al. (2018) implement such a platform for exploring the results of their topic analysis of the Greek parliament. Similar initiatives could be developed for the ParlaMint datasets in order to facilitate cross-national comparison and research.

BIBLIOGRAPHY

Abercrombie, G. (2021). *Topic-Centric Sentiment Analysis of UK Parliamentary Debates* [Ph.D., The University of Manchester (United Kingdom)].

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

Abercrombie, G., & Batista-Navarro, R. (2018, May). `Aye' or `No'? Speech-level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. LREC 2018, Miyazaki, Japan.

Banducci, S. A., Gidengil, E., & Everitt, J. (2012). Women as political communicators. Candidates and campaigns. *The Sage Handbook of Political Communication*, 164–185.

Bhatia, S., & P, D. (2018). Topic-Specific Sentiment Analysis Can Help Identify Political Ideology. *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 79–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bor, D., Lee, B. S., & Oughton, E. J. (2023). *Quantifying polarization across political groups on key policy issues using sentiment analysis* (arXiv:2302.07775). arXiv.

Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*.

Brescoll, V. L., & Uhlmann, E. L. (2008). Can an angry woman get ahead? Status conferral, gender, and expression of emotion in the workplace. *Psychological Science*, *19*(3), 268–275.

Chen, Y., Peng, Z., Kim, S.-H., & Choi, C. W. (2023). What We Can Do and Cannot Do with Topic Modeling: A Systematic Review. *Communication Methods and Measures*, *0*(0), 1–20.

Chen, Y., & Skiena, S. (2014). Building Sentiment Lexicons for All Major Languages. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 383–389.

Christiansen, F. J. (2017). Vendes karrieren på hovedet?: Anciennitets betydning for om folketingsmedlemmer opnår politiske lederposter. *Oekonomi Og Politik*, *90*(3), 27–39.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

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

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 & Linguistic Computing*, *27*(2), 139–153.

Damgaard, E. (2004). Developments in Danish Parliamentary Democracy: Accountability, Parties and External Constraints. *Scandinavian Political Studies*, *27*(2), 115–131.

De Giorgi, E., & Marangoni, F. (2015). Government laws and the opposition parties' behaviour in parliament. *Acta Politica*, *50*(1), 64–81.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*.

Diermeier, D., Godbout, J.-F., Yu, B., & Kaufmann, S. (2012). Language and Ideology in Congress. *British Journal of Political Science*, *42*(1), 31–55.

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(7).

Elmelund-Præstekær, C., & Schumacher, G. (2014). Én for alle og alle for én? *Politica*, *46*(3), Article 3.

Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Grigorova, V., Rudolf, M., Pančur, A., Kopp, M., Barkarson, S., Steingrímsson, S. hór, van der Pol, H., Depoorter, G., de Does, J., Jongejan, B., Haltrup Hansen, D., Navarretta, C., Calzada Pérez, M., de Macedo, L. D., … Rayson, P. (2021). *Multilingual comparable corpora of parliamentary debates ParlaMint 2.1*. http://hdl.handle.net/11356/1432

Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Pančur, A., Rudolf, M., Kopp, M., Barkarson, S., & Steingrímsson, S. (2022). The ParlaMint corpora of parliamentary proceedings. *Language Resources and Evaluation*, 1–34.

Fellbaum, C. (2010). WordNet. In R. Poli, M. Healy, & A. Kameas (Eds.), *Theory and Applications of Ontology: Computer Applications* (pp. 231–243). Springer Netherlands.

Fiva, J. H., Nedregård, O., & Øien, H. (2021). *Polarization in Parliamentary Speach* (SSRN Scholarly Paper No. 3767690).

Folketinget. (2017, January 9). *Hvilke regler er der for sprog og opførsel i Folketingssalen?* Folketinget. https://www.ft.dk/da/ofte-stillede-spoergsmaal/mode_hvilke-regler-er-der-for-sprog-og-opfoersel-i-folketingssalen

Folketinget. (2023). *Kvinder i Folketinget*. Folketinget. https://www.ft.dk/da/folkestyret/grundloven-og-folkestyret/kvinder-i-politik#30E2BC07610C4259AF2AEA350FB395C7

Gentzkow, M., Shapiro, J. M., & Taddy, M. (2016). *Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech* (No. w22423). National Bureau of Economic Research.

Gkoumas, D., Pontiki, M., Papanikolaou, K., & Papageorgiou, H. (2018). Exploring the Political Agenda of the Greek Parliament Plenary Sessions. *Proceedings of the LREC 2018 Workshop "ParlaCLARIN: Creating and Using Parliamentary Corpora*.

Glick, P., & Fiske, S. T. (1997). Hostile and Benevolent Sexism. *Psychology of Women Quarterly*, *21*(1), 119–135.

Goet, N. D. (2019). Measuring polarization with text analysis: Evidence from the UK House of Commons, 1811–2015. *Political Analysis*, *27*(4), 518–539.

Green-Pedersen, C., & Skjæveland, A. (2020). Blokpolitik og nye politiske emner. Hvordan går det med samarbejdsmønstrene i Folketinget? *Politica*, *52*(3), Article 3.

Gregersen, S. (2019). Årsagssætninger med fordi i (formelt) dansk talesprog. *NyS, Nydanske Sprogstudier*, *57*.

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 267–297.

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (arXiv:2203.05794). arXiv.

Han, S. (2022). Elite polarization in South Korea: Evidence from a natural language processing model. *Journal of East Asian Studies*, *22*(1), 45–75.

Hansen, D. H., Navarretta, C., & Offersgaard, L. (2018, May 7). A Pilot Gender Study of the Danish Parliament Corpus. *Proceedings of the LREC 2018 Workshop "ParlaCLARIN: Creating and Using Parliamentary Corpora", Darja Fišer, Maria Eskevich, Franciska de Jong (Eds.)*. Miyazaki, Japan.

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

Hargrave, L., & Blumenau, J. (2022). No Longer Conforming to Stereotypes? Gender, Political Style and Parliamentary Debate in the UK. *British Journal of Political Science*, *52*(4), 1584–1601.

Hargrave, L., & Langengen, T. (2021). The gendered debate: Do men and women communicate differently in the house of commons? *Politics & Gender*, *17*(4), 580–606.

Haselmayer, M., Dingler, S. C., & Jenny, M. (2022). How Women Shape Negativity in Parliamentary Speeches—A Sentiment Analysis of Debates in the Austrian Parliament. *Parliamentary Affairs*, *75*(4), 867–886.

Hirst, G., Riabinin, Y., & Graham, J. (2010). Party status as a confound in the automatic classification of political speech by ideology. *Proceedings of the 10th International Conference on Statistical Analysis of Textual Data (JADT 2010)*, 731–742.

Hjorth, E., & Kristiansen, K. (2005). *Den Danske Ordbog (DDO)*. Det Danske Sprog- og Litteraturselskab & Gyldendal. https://ordnet.dk/ddo

Holm, L. (2006). *Folketinget og Udlændingepolitikken-diskurser om naturaliserede, indvandrere og flygtninge 1973-2002*. Institut for Historie, Internationale Studier og Samfundsforhold, Aalborg Universitet. Spirit PhD Series Nr. 4

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1.

Ilie, C. (2015). Parliamentary Discourse. In *The International Encyclopedia of language and social interaction*. John Wiley & Sons.

Kapukaranov, B., & Nakov, P. (2015). Fine-Grained Sentiment Analysis for Movie Reviews in Bulgarian. *Proceedings of the International Conference Recent Advances in Natural Language Processing*, 266–274.

Karasimeonov, G. (2019). *Partiynata sistema v Balgariya [The party system in Bulgaria]*. ГорексПрес.

Koeva, S. (2021). The Bulgarian WordNet: Structure and specific features. *Papers of BAS Humanities and Social Sciences*, *8*(1), 47–70.

Kolarova, R. (2018). Tri izmereniya na balgarskiya parlamentarizam 1991-2018 [Three Dimensions of Bulgarian Parliamentarism 1991-2018]. *Политически изследвания*, *1–2*, 5–40.

Koppensteiner, M., Stephan, P., & Jäschke, J. P. M. (2016). Moving speeches: Dominance, trustworthiness and competence in body motion. *Personality and Individual Differences*, *94*, 101–106.

Lauridsen, G. A., Dalsgaard, J. A., & Svendsen, L. K. B. (2019). SENTIDA: A New Tool for Sentiment Analysis in Danish. *Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift*, *4*(1), Article 1.

Laver, M., Benoit, K., & Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review*, *97*(2), 311–331.

Ljubešić, N., & Dobrovoljc, K. (2019). What does Neural Bring? Analysing Improvements in Morphosyntactic Annotation and Lemmatisation of Slovenian, Croatian and Serbian. *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, 29–34. https://doi.org/10.18653/v1/W19-3704

Mandravickaitė, J., & Oakes, M. (2016). Multiword Expressions for Capturing Stylistic Variation Between Genders in the Lithuanian Parliament. *CCURL 2016 Collaboration and Computing for Under-Resourced Languages: Towards an Alliance for Digital Language Diversity*, *10727*(2357596), 80.

Mavrodieva, I. (2014). *The concept of the Bulgarian "European Identity" across The Bulgarian Language and Political Rhetoric in Bulgaria*. *12*, 1–16.

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61).

Meden, K. (2022). Speech-level Sentiment Analysis of Parliamentary Debates using Lexicon-based Approaches. *Conference on Language Technologies and Digital Humanities*. Ljublana, 2022

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113.

Miok, K., Hidalgo-Tenorio, E., Osenova, P., Benitez-Castro, M.-A., & Robnik-Sikonja, M. (2022). *Multi-aspect Multilingual and Cross-lingual Parliamentary Speech Analysis* (arXiv:2207.01054). arXiv.

Mochtak, M., Rupnik, P., & Ljubešič, N. (2022). *The ParlaSent-BCS dataset of sentiment-annotated parliamentary debates from Bosnia-Herzegovina, Croatia, and Serbia* (arXiv:2206.00929).

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2017). Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis*, *16*(4), 372–403.

Müller-Hansen, F., Callaghan, M. W., Lee, Y. T., Leipprand, A., Flachsland, C., & Minx, J. C. (2021). Who cares about coal? Analyzing 70 years of German parliamentary debates on coal with dynamic topic modeling. *Energy Research & Social Science*, *72*, 101869.

Navarretta, C., Haltrup Hansen, D., & Jongejan, B. (2022). Immigration in the Manifestos and Parliament Speeches of Danish Left and Right Wing Parties

between 2009 and 2020. *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, 71–80.
Navarretta, C., & Hansen, D. H. (2020). Identifying Parties in Manifestos and Parliament Speeches. *Proceedings of the Second ParlaCLARIN Workshop*, 51–57.
Navarretta, C., & Hansen, D. H. (2022). The Subject Annotations of the Danish Parliament Corpus (2009-2017)—Evaluated with Automatic Multi-label Classification. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 1428–1436.
Nielsen, F. Å. (2018). *Danish resources* [Data set]. https://bit.ly/2NDHcbW
Nimb, S., Olsen, S., Pedersen, B., & Troelsgård, T. (2022). A Thesaurus-based Sentiment Lexicon for Danish: The Danish Sentiment Lexicon. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2826–2832.
Onyimadu, O., Nakata, K., Wilson, T., Macken, D., & Liu, K. (2014). Towards Sentiment Analysis on Parliamentary Debates in Hansard. In W. Kim, Y. Ding, & H.-G. Kim (Eds.), *Semantic Technology* (pp. 48–50). Springer International Publishing.
Osenova, P. (2022a). Raising and Control Constructions in a Bulgarian UD Parsebank of Parliament Sessions. *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, 68–74.
Osenova, P. (2022b). THE COVID-19 PANDEMIC IN THE VALENCY OF ITS PREDICATES: OBSERVATIONS ON A CONTEMPORARY CORPUS OF PARLIAMENTARY DEBATES. *Journal of Bulgarian Language*, *69*(PR), 113–121.
Pedersen, B. S., & Asmussen, J. (2006). DanNet–fra ordbog til et leksikalsksemantisk WordNet for dansk. *Leda-Nyt*, *42*, 3–12.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.
Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.
Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). *Deep contextualized word representations* (arXiv:1802.05365). arXiv. https://doi.org/10.48550/arXiv.1802.05365

Petersen, M. N. (2009). Fra barnets tarv til ligestilling-en queerteoretisk undersøgelse af Folketingets forhandlinger om kunstig befrugtning. *Kvinder, Køn & Forskning*, *2*.

Peterson, A., & Spirling, A. (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems. *Political Analysis*, *26*(1), 120–128.

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

Proksch, S.-O., & Slapin, J. B. (2012). Institutional Foundations of Legislative Speech. *American Journal of Political Science*, *56*(3), 520–537.

Rheault, L., Beelen, K., Cochrane, C., & Hirst, G. (2016). Measuring Emotion in Parliamentary Debates with Automated Textual Analysis. *PLOS ONE*, *11*(12), e0168843.

Richardson, L. (2007). Beautiful soup documentation. *April*.

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

Sakamoto, T., & Takikawa, H. (2017). Cross-national measurement of polarization in political discourse: Analyzing floor debate in the U.S. the Japanese legislatures. *2017 IEEE International Conference on Big Data (Big Data)*, 3104–3110.

Simov, K. I., Osenova, P., Slavcheva, M., Kolkovska, S., Balabanova, E., Doikoff, D., Ivanova, K., Simov, A., & Kouylekov, M. (2002). Building a Linguistically Interpreted Corpus of Bulgarian: The BulTreeBank. *LREC*.

Skubic, J., & Fišer, D. (2022). Parliamentary Discourse Research in Sociology: Literature Review. *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, 81–91.

Slapin, J. B., & Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, *52*(3), 705–722.

Søyland, M. G., & Lapponi, E. (2017). Party polarization and parliamentary speech. *Working Paper*.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). *The general inquirer: A computer approach to content analysis* (p. 651). M.I.T. Press.

Tanasoiu, C., & Racovita, M. (2012). Post-Accession (Anti-)Corruption Record in Romania and Bulgaria. *L'Europe en Formation*, *364*(2), 243–263.

Tang, D., Wei, F., Qin, B., Yang, N., Liu, T., & Zhou, M. (2016). Sentiment Embeddings with Applications to Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, *28*(2), 496–509.

Tarasheva, E. (2004). Osobenosti na parlamentarniya diskurs: Kakvi danni dava ezikat za sotsialnoto yavlenie? [Peculiarities of parliamentary discourse: What data does language provide about the social phenomenon?]. *Език, Литература, Култура*, 26–35.

Tarasheva, E. (2014). Osobenosti na parlamentarniya diskurs–danni ot korpusno izsledvane na parlamentarnite protokoli na 42to Narodno sabranie [Peculiarities of parliamentary discourse—Data froma corpus study of the parliamentary proceedings of the 42nd National Assembly]. In: Научна конференция "Хуманитарните науки днес", 2014, Читалище "Възраждане", Sofia, Bulgaria

TEI Consortium (Ed.). (2023). *TEI P5: Guidelines for Electronic Text Encoding and Interchange. [4.6.0]*. TEI Consortium.

Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *Acm Transactions on Information Systems (Tois)*, *21*(4), 315–346.

Viera, A. J., & Garrett, J. M. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*, *37*(5), 360–363.

Vries, E. de. (2022). The Sentiment is in the Details: A Language-agnostic Approach to Dictionary Expansion and Sentence-level Sentiment Analysis in News Media. *Computational Communication Research*, *4*(2), 424–462.

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021.

Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G. H., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2019). *Multilingual Universal Sentence Encoder for Semantic Retrieval* (arXiv:1907.04307). arXiv.

Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, *29*(2), 205–231.

Yu, B., Kaufmann, S., & Diermeier, D. (2008). Classifying party affiliation from political speech. *Journal of Information Technology & Politics*, *5*(1), 33–48.

APPENDIX

Appendix A. Political Parties

Denmark	
Left	Socialdemokraterne (S) Radikale Venstre (RV) Socialistisk Folkeparti (SF) Enhedslisten (EL) Alternativet (ALT)
Right	Venstre (V) Konservativ Folkeparti (KF) Liberal Alliance (LA) Dansk Folkeparti (DF) Nye Borgerlige (NB)
Bulgaria	
Left	Bulgarian Socialist Party – Left Bulgaria (BSPLB) Alternative for Bulgarian Revival (ABV) We Continue the Change (WCC) Bulgarian Socialist Party – for Bulgaria (BSPFB) Rise Up! Thugs Out! (RUTO)
Right	Reformist Bloc (RB) Patriotic Front (PF) Bulgarian Democratic Centre – National Union (BDC-NU) Ataka Party (AP) GERB Party (GERB) GERB-Union of Democratic Forces (GERB-UDF) Movement for Rights and Freedoms (MRF) There Is Such a People (TISP) Revival Party (RP) Democratic Bulgaria (DB) United Patriots (UP) Volya (VOLYA)

Table A.1 Left-wing and right-wing parties active in the Danish and Bulgarian parliaments in the years 2014-2022.

Appendix B. Annotator Guidelines (English Translation)

Excerpts are considered positive when they:

- Express positive feelings and attitudes (e.g., joy, hope, kindness, praise)
- Use words with positive charge and connotation (e.g., *success*, *safe*, *achievement*) without negation
- Emphasize the positive aspects of the subject/issue and discuss them in a predominantly positive and optimistic light (e.g., “we must improve and secure the roads for the benefit of everyone” rather than “we need to repair all the bad and dangerous roads”)

Excerpts are considered negative when they:

- Express negative feelings and moods (e.g., anger, hatred, hostility)
- Use words with negative charge and connotation (e.g., *error*, *crooked*, *unmanageable*) without negation
- Constitute explicit insults, accusations, or attacks
- Emphasize the negative aspects of the subject/issue and discuss them in a predominantly negative and pessimistic light (e.g., “we need to repair all the bad and dangerous roads” instead of “we must improve and secure the roads for the benefit of everyone”)

The focus is on the valence of the language and not so much specific positions, although these often overlap. “How”, rather than “what”, is said. So if an expression of support (e.g., "SF supports the proposal") is not followed up by other positive expressions, feelings, etc., it is considered neutral. If an expression of support is followed by criticism with a negative charge (e.g., "While SF supports the proposal, the problem is that..."), it is negative despite the support. And vice versa.

Critical questions ("How will it be financed?") are considered neutral if negative words are not explicitly used, etc. ”Thank you” at the start of the statement is not considered positive, as it is just procedure.

The excerpts are measured on the following scale:

- -2 strongly negative
- -1 predominantly negative
- 0 neutral
- 1 predominantly positive
- 2 strongly positive

Neutral excerpts are both those that do not express emotions/subjectivity at all, as well as those that have an equal amount for positivity and negativity and are therefore difficult to categorize.

In general, it will often be difficult to decide on a rating, so do it according to intuition or first impressions when unsure. Sentiment is a fuzzy concept.

Appendix C. ParlaMint-DK Additional Statistics

Party Status	MPs (%)	Speakers (%)	Words (%)	Speeches (%)	Avg. Speech Length
Coalition	28,0	32,3	32,7	30,6	215,4
Minister	10,6	12,2	21,0	18,1	233,7
Regular MP	17,4	20,1	11,7	12,5	188,9
Opposition	47,9	42,4	46,3	46,3	197,6
Other	24,1	25,3	21,0	22,1	191,4

Table C.1 Percentage of speakers, words, speeches, and average speech length for speakers by party and minister status in the Danish parliament 2014-2022.

Bloc	F MPs (%)	F Words (%)	F Speeches (%)	F Speech Length	M Speech Length
Left	43,6	37,3	37,0	202,0	199,6
Right	35,2	29,5	30,2	184,2	190,8
Other	36,4	25,1	25,5	168,5	172,3

Table C.2 Percentage of female (F) speakers, words, and speeches by left-right bloc, as well as a comparison of male (M) and female speech length per bloc in ParlaMint-DK. Left and right categories are assigned based on Tab. A.1

Proceeding Type	Party Status	Words (%)	Speeches (%)	Avg. Speech Len.
Reading	Coalition	28,6	26,7	225,9
	Opposition	49,1	49,7	208,3
	None	22,2	23,5	198,8
Negotiation	Coalition	30,9	26,8	250,4
	Opposition	45,1	46,9	205,6
	None	24,0	26,8	190,8
Question	Coalition	55,9	50,7	170,4
	Opposition	34,5	38,7	137,9
	None	9,5	10,6	139,9

Table C.3 Percentage of words and speeches, as well as average speech length by proceeding type and party status in ParlaMint-DK.

Party Status	MPs (%)	Speakers (%)	Speeches (%)	Words (%)	Avg. Speech Length
Coalition	53,0	59,6	47,0	49,2	209,6
Opposition	43,5	45,5	48,7	46,9	194,5
Other	8,2	8,0	3,8	3,8	203,7

Table D.1 Percentage of speakers, words, speeches, and average speech length for speakers by party status in the Bulgaria parliament 2014-2022. The percentage of MPs does not add up to 100 because of changes in party status.

Bloc	F MPs (%)	F Words (%)	F Speeches (%)	F Speech Length	M Speech Length
Left	21,8	19,2	17,9	195,0	178,0
Right	26,0	29,2	27,8	218,2	203,5
Other	11,1	0,2	0,5	72,7	177,6

Table D.2 Percentage of female (F) speakers, words, and speeches by left-right bloc, as well as a comparison of male (M) and female speech length per bloc in ParlaMint-DK. Left and right categories are assigned based on Tab. A.1

Debate Type	Party Status	Words (%)	Speeches (%)	Avg. Speech Len.
Voting/Reading	Coalition	54,8	53,4	212,4
	Opposition	45,1	46,7	200,3
	None	3,3	3,1	218,8
Draft Decision/ Procedural	Coalition	46,8	43,9	207,3
	Opposition	53,1	56,0	184,7
	None	5,9	6,4	179,2
Hearing/Control	Coalition	41,4	39,6	182,1
	Opposition	58,6	60,4	168,8
	None	6,4	7,6	146,6

Table D.3 Percentage of words and speeches, as well as average speech length by proceeding type and party status in ParlaMint-BG.

Appendix E. Excluded Words

Danish				
Word	**Translation.**	**Polarity**	**Crit.**	**Explanation**
opposition (n.)	opposition (n).	neg	2	Political topic/institution in parliament.
skat (n.)	dear (n.), tax (n.)	pos	1, 2	Polysemy with ambiguous orientation, political topic
velfærd (n.)	welfare (n.)	pos	2	Political topic/technical meaning
blasfemi (n.)	blasphemy (n.)	neg	2, 3	Unusual word, also political topic in recent years (REF)
lide (v.)	suffer (v.)	neg	1	A kind of polysemy, as it means "to like" in the common compound phrase "kan lide".
selvsikker (adj.)	self-confident (adj.)	pos	1	Can also have negative connotations, as in "self-absorbed" (listed as synonym in DDO).
udmærket (adj.)	excellent (adj.), acceptable (adj.)	pos	1	Polysemy with ambiguous orientation
bæredygtig (adj.)	sustainable (adj.)	pos	2	Political topic (environment)
lede (n.)	disgust (n.)	neg	1, 3	As a verb it means "to lead", which does not have a negative connotation, also a rare word
deprimeret (adj.)	depressed (adj.)	neg	2	As all words referring to and naming sickness/health, it is a potential political topic, e.g., youth mental health
Bulgarian				
Word	**Translation**	**Polarity**	**Crit.**	**Explanation**

действащ (adj.)	active (adj.)	pos	1	Polysemy, can also mean “current”, as in “the current pope”
съюзник (n.)	ally (n.)	pos	2	Political topic
жертва (n.)	sacrifice (n.)	neg	1	Can have a positive connotation in many cases
корупция (n.)	corruption (n.)	neg	2	Major political topic
несъстоятелен (adj.)	insolvent (adj.), unacceptable (adj.)	neg	1,2	Polysemy with negative polarity in both cases, but possible political topic
геноцид (n.)	genocide (n.)	neg	2,3	Possible political topic and unusual in colloquial speech

Table E.1 Sample of words removed from seed lists based on inclusion criteria with explanations.

Appendix F. Top Expansion Words

Danish		Bulgarian	
Positive	Negative	Positive	Negative
takke *thank*	forårsage *cause*	осигуря *ensure*	причиниявам *inflict*
samarbejde (n.) *collaboration*	utilsigtet *unintentional*	гарантирам *guarantee (v.).*	очевиден *obvious*
glæde *gladden*	eskalere *escalate*	заедно *together*	закононарушение *breach of law*
repræsentation *representation*	spekulation *speculation*	подкрепа *support*	пораждам *give rise to*
levere *deliver*	afværge *avert*	пожелая *wish, (v.)*	инфлационен *inflationary*
flot *nice, beautiful*	fyring *firing*	нека *let*	предизвикам *provoke*
kvittere *acknowledge*	komplikation *complication*	гарантиране *guaranteeing (n.)*	вследствие *as a consequence*
opbakning *support*	hærværk *vandalism*	усилие *effort*	безхаберие *carelessness*
dejlig *wonderful*	fatal *fatal*	разчитам *count on*	повред *harm*
tilbyde *offer (v.)*	oversvømmelse *flooding*	давам *give*	авария *accident*
sikker *certain*	teknikalitet *technicality*	приоритет *priority*	умишлено *deliberately*
tillykke *congratulations*	overfald *assault*	запазя *keep*	пораждам (се) *give rise to*
kvalificere *qualify*	unødig *unnecessary*	осигурявам *ensure*	форсмажорен *force majeure*
samarbejde (v.) *collaborate*	forældelse *obsolescence*	намеря *find*	констатирам *assert*

Table F.1 Top 10 most positive and negative expansion words for both languages, based on vector similarity to seed lists.