

Parliamentary Discourse Research in Political Science: Literature Review

Jure Skubic, Darja Fišer

Institute of Contemporary History

Privoz 11, 1000 Ljubljana

jure.skubic@inz.si, darja.fiser@inz.si

Abstract

One of the major research interests for political science has always been the study of political discourse and parliamentary debates. This literature review offers an overview of the most prominent research methods used in political science when studying political discourse. We identify the commonalities and the differences of the political science and corpus-driven approaches and show how parliamentary corpora and corpus-based approaches could be successfully integrated in political science research.

Keywords: parliamentary discourse, political science, parliamentary corpora

1. Introduction

Parliamentary debates are one of the best sources of information about political discourse, which is inherently valuable for research in the humanities and social sciences. Especially political science is particularly involved in the analysis of political power and authority exercised through parliamentary discourse.

This literature review is part of a series of literature reviews produced as part of the ParlaMint project (Erjavec et al., 2022). Similar reviews have been compiled for sociology (Skubic and Fišer, 2022) and history (Skubic and Fišer, 2022) and are important for better understanding how the humanities and social sciences use qualitative and quantitative research methods in analyzing parliamentary discourse. The ParlaMint project has developed comparable corpora of parliamentary transcripts for more than 20 European countries and offered literature reviews, showcases, and tutorials mentioned earlier to promote the use of the corpora in a wide range of scholarly communities interested in the study of parliamentary discourse and debate. In this paper, we review existing political science research focusing on written parliamentary records and the commonly used research methods. We view these approaches as complementary to other common political science research techniques and types of data sources such as surveys, records of election results, media content, etc.

This literature review is organized as follows. In the first part, we outline the selection process of relevant papers and explain the research methods they employ. In the second part, we summarize each of the selected papers in terms of 1) the research topic, 2) the data collection, 3) the research method, and 4) a brief discussion of possible improvements to the research. We conclude the review with a discussion of how this area of political science could benefit from the use of corpus data and the use of corpus-assisted research methods or other text mining methods.

2. Political Science Methods

Parliamentary discourse is an important focus of political science research at the (inter)national or local level. Like many other social sciences, it draws on and complements various methodological traditions in the study of politics and governance, legislation, and political discourse to increase the relevance and reliability of its research findings (Lauer, 2021). The methodological pluralism of political science allows it to address contemporary issues and problems that arise in the broad field of social sciences in general (Franco et al., 2021), and to focus on topics that might go unaddressed in other social science disciplines. Although political science has in recent years taken a quantitative turn (ibid.), qualitative methods and approaches are still widely used, prove to be highly effective, and provide meaningful insights into important research questions.

Blaxill (2022) notes that political scientists are interested in language and discourse as a means of studying political power, change, institutions, etc. Since political discourse is about the text and speech of professional politicians and political institutions (van Dijk, 1997), documentary sources are a valuable source of data for political science. Documents (texts, laws, etc.) are usually collected from official websites or archives of relevant organizations (parliaments, libraries, etc.) or by visiting archives, bureaus, and other organizations (Franco, 2021). In addition, political scientists often triangulate data and metadata collected from official parliamentary minutes and policy texts with data from other sources such as interviews, (social) media, newspapers, etc. This makes the ParlaMint corpora directly relevant for political science researchers.

3. Literature Selection

3.1 Selection of Papers

When selecting relevant papers for this literature review, the following criteria were followed. We used the following scholarly search engines to search for relevant papers:

- Elsevier (<https://www.elsevier.com>)
- Project MUSE (<https://muse.jhu.edu>)

- SAGE Journals (<https://journals.sagepub.com>)
- Springer Link (<https://link.springer.com>)
- Taylor and Francis Online (<https://www.tandfonline.com>)
- Wiley Online Library (<https://onlinelibrary.wiley.com>)

We searched for keywords such as “parliamentary discourse”, “parliamentary debates” and “parliamentary proceedings” and applied the following filters to narrow down search options:

- Publication period: 2012 – 2022,
- Discipline: political science,
- Article ranking: “most relevant” or “most cited”,
- Relevant journals: additional filters were sometimes needed to search papers in relevant journals.

Because the number of papers was still high, we performed an additional selection process, analyzing the abstract, topic, data collection, and methods used for each paper. At this stage, many papers were screened out either because of a lack of methodological explanation or because the research did not focus on parliamentary data. We focused only on papers that specifically addressed parliamentary and/or legislative documents or interwove them with other data sources. After completing the selection process, we selected 24 relevant papers from the following political science journals: Parliamentary Affairs, British Politics, French Politics, Comparative European Politics, Political Communication, Ethnicities, The British Journal of Politics and International Relations, Australian Journal of Political Science, Political Analysis Journal, and Journal of Contemporary European Studies.

3.2 Overview of Methods

All 24 papers relevant for this review are listed in the Google Spreadsheet.¹ We thoroughly analyzed all of them and paid special attention not only to the data selection process or the methods employed, but also to the thematic focus of each paper. In all the reviewed papers, methods and data selection process were clearly explained and all of them used parliamentary records as the main source of data. The research questions of the analyzed papers were highly heterogeneous, so we decided not to group the papers thematically. Out of 24 analyzed papers, 12 employed content analysis, 3 (critical) discourse analysis, 2 sentiment analysis, 2 thematic analysis, 2 papers employed a mixed methods approach and 3 papers employed one of the many text-as-data approaches (1 paper social network analysis and 2 papers quantitative text analysis with supervised machine learning techniques). Due to methodological and in a few cases thematic similarities of some

papers, we decided not to include all 24 papers in this review but analyze no more than 2 representative papers for each methodological approach.

4. Reviewed Research and Methods

4.1 Content Analysis

Content analysis (CA) is one of the most widely used research techniques in social sciences and its main goal is to analyze data in a specific context and extract meaningful information from the analyzed documents (Krippendorff, 2018). According to Blassnig (2022), it is perhaps one of the most used methods in the field of political science, mainly due to the general influence of other disciplines such as sociology, history, philosophy, etc. It is used to interpret textual data through the process of coding and identifying themes or patterns (Lilja, 2021), and to analyze the self-representation of political actors through the analysis of political and parliamentary speeches, debates, party platforms, etc. In political science, researchers often decide for triangulation of content analysis with other either qualitative (e.g., CDA) or quantitative (often digital) methods. Due to advances in computational research approaches, content analysis is becoming more and more digitized with researchers using computer software to systematically import and analyze large volume of text documents without spending considerable amount of time reading or paying for expensive coding (Provalis research, 2019). Although qualitative content analysis still prevails in political science, quantitative content analysis is once again gaining recognition and is becoming increasingly more popular.

4.1.1 Emotions in EU Parliamentary Debates

Research problem: The aim of Sanchez Salgado’s (2021) paper was to explore the verbal display and role of emotions in the European Parliament (EP). She analyzed how emotions are expressed inside the EP and how they reflect not only power but also status dynamics.

Data collection: The data for her research consisted of 25 plenary debates in English, French, Dutch and Spanish that took place in EP between 2009 and 2017. The author focused on two topics in EP in which she expected emotions to play a crucial role: the financial crisis (2009 – 2014, 14 debates) and the refugee crisis (2014 – 2017, 11 debates) as the two most challenging crises that the EU had faced before 2020. She accessed the debates on the website of the EP in September 2017. For the first she selected those debates which included “economic crisis” or “financial crisis” in their title as for the latter she used the debates the title of which included the words “migration” and “refugees”. The automatic coding she employed only included keywords which correspond to basic primary and secondary emotions as defined by Parrot (2001). She was particularly interested in analysis of emotional patterns and structures which

¹ Those papers can be found in the first sheet in the Google Spreadsheet titled “All papers”. The second sheet, titled “Papers selected for report” includes papers, which

are in detail discussed below. Link to the spreadsheet: <https://docs.google.com/spreadsheets/d/1dd9YCDs9G53NBxxg0Bxhfbjx3QxjWjgN4tPjCLw2VWg/edit#gid=0>.

were identified through an in-depth reading of all the debates in which emotion markers were used within their discursive context.

Research method: The author opted for an in-depth (qualitative) content analysis of 25 debates in EP in which she observed explicit emotion keywords present in discourses. For qualitative content analysis of emotions, she used the Atlas.ti² data analysis software (ATLAS.ti Scientific Software Development GmbH, 1993), which contributed to the efficiency, consistency, and transparency of her analysis. In her Atlas.ti analysis she considered only emotion keywords, whereas the in-depth contextual analysis accounted for all types of implicit and explicit references to emotions.

Discussion: Sanchez Salgado's research is one in a few which focuses on international (EU) parliamentary debates. What could be seen as a potential shortcoming of the research is in author not elaborating on why she specifically chose debates in those languages and not any other. She points out that the transcriptions since 2012 are not available in English, which could be seen as a limitation, however it also shows how emotions are expressed in various languages.

4.1.2 Exploring Feminist Arguments in German Parliamentary Debates

Research problem: Och (2019) analyzed the parliamentary discourse around two instances of feminist policy adoption in two conservative German governments. She showed that in both analyzed governments feminist arguments dominated the debates.

Data collection: Och analyzed documents from the 16th (2006) and 18th (2015) legislative period of the German parliament. She identified suitable documents with the help of document and information system for parliamentary processes of the Bundestag. This system returned all parliamentary documents linked to respective bills, which included verbatim protocols of plenary debates in both chambers as well as verbatim protocols of the committee hearings and bill documents presented by the federal government to parliament for information purposes or in response to parliamentary questions. She also included documents published by Federal Ministries for Family Affairs, Senior Citizens, Women and Youth, Federal Ministry of Justice and Consumer Affairs as well as statements and speeches by the responsible ministers if they directly discussed the respective policy.

Research method: The author employed qualitative content analysis on a series of parliamentary documents of the German parliament by reading all the documents and coding them by hand to identify statements that contained arguments of either of the two broad coding categories: utility-driven arguments and feminist arguments. She coded arguments as utility-driven if the policy was justified as a means to a

non-feminist end and as feminist if they showed feminist attitudes and behavior (referring to gender equality, sex-based discrimination, inequalities or challenging the elimination of traditional gender roles) as defined by Carroll (1984).

Discussion: Och was the only coder and coded all the texts by hand. This could be identified as a potential research problem which could be avoided if more coders were involved in coding process and if computer-assisted methods were used to avoid coding by hand.

4.2 Discourse Studies

Discourse Studies has been developing at the intersection of language and society. It combines various qualitative and quantitative research methods as well as different genres such as news reports and parliamentary debates (van Dijk, 2018). In this review, we identified two salient methods of Discourse Studies, namely discourse analysis (DA) and critical discourse analysis (CDA).

In political science, **discourse analysis (DA)** is most frequently used to study parliamentary debates and parliamentary discourse. It is frequently referred to as political discourse analysis (PDA) (Dunmire, 2012) and can sometimes be mistakenly equated to content analysis even though it does not focus on the analysis of content but rather on the analysis of language through specific text and context. One of the main foci of DA is to examine how political power, power abuse and domination manifest through discourse practices and structures (ibid.).

Critical discourse analysis (CDA), or critical-political discourse analysis, is one of the most visible categories of discourse studies frequently applied to parliamentary communication. It provides a critical context in which political debates occur and analyzes the relationship between power and the traditional ideology in implied discourse (van Dijk, 2018). A contribution that CDA can make to political studies is mostly in offering a general theoretical perspective on discourse which recognizes the constitutive potential of discourse within and across social practices without reducing social practices to their discursive aspect (Farrelly, 2010).

4.2.1 Parliamentary Discourse on Immigration

Research problem: May (2016) analyzed the parliamentary discourses on immigration in Canada and France and wanted to find out what arguments were introduced in parliamentary arenas to justify more restrictive immigration policies.

Data collection: May's analysis was stretched between January 2006 to December 2013. The two countries were chosen because of the very similar discussions about immigration and because they developed different models of integration and management of cultural diversity. He analyzed parliamentary debates following seven bills which included a high number of immigration indicators.

² <https://atlasti.com>

During the coding procedure he and another coder read through the debates and compiled a list of coding units which was inspired by the literature review. Then they identified the phrases and clusters of meaning which resulted in a hierarchical coding structure which included 32 nodes. They refined the coding procedure by introducing new nodes based on the themes they considered relevant, which resulted in the introduction of new nodes into the structure. After that the inter-coder reliability test was performed followed by the discursive analysis.

Research method: The author employed critical discourse analysis (CDA). After identifying the main 32 keywords (refugee, asylum seekers, Roma, financial cost, immigration, multiculturalism, etc.) in the chosen parliamentary debates, he opted for lexical analysis with the Nvivo software³ (QSR International Pty Ltd., 2020) to code specific discursive constructions.

Discussion: May gave no specific account as to where the analyzed debates were downloaded from and what language they were in (relevant for Canada which is bilingual). The paper could also benefit from a more thorough description of the discourse analysis since it is mentioned as the primary method used.

4.2.2 Political Discourse about COVID-19

Research problem: Jarvis (2021) analyzed the conceptions of time during the COVID-19 pandemic within the UK parliamentary discourse. He showed that construction of temporality was important for social, political, and historical positioning of the virus and that such constructions had impact on UK government's response to the virus.

Data collection: Jarvis analyzed more than 120 texts including parliamentary speeches, newspaper articles, press releases, public letters, accouchements, and policy statements. The timeframe of his analysis was limited to the first six months of 2020 since this was the timeframe crucial to the government's communication of the crisis. He designed his own corpus by collecting the texts directly from the official website of the Prime Minister's office. All the texts were thoroughly read to determine their relevance for the research and all the texts that referred to the pandemic or its response were included in the corpus for future analysis. Jarvis organized coding material around various index categories (the virus, the UK government's response, the scientific response, the public, temporality) and reread all the texts through his framework. This allowed for the distribution and coding of the data according to different themes and their subcategories.

Research method: The author employed discourse analysis via the framework method as defined by Ritchie and Spencer (2002). He analyzed qualitative data through summarizing, sifting, and sorting research material and classifying large volumes of data in its own terms. Jarvis performed a detailed analysis which involved a thorough reading of the

corpus in four stages: 1) familiarization with the documents, 2) coding via paraphrasing of short text sections, 3) developing an analytical framework from the coded material, and 4) applying this framework to the corpus.

Discussion: Jarvis' paper shows the importance of collecting data from various sources and strengthens the notion that political scientists often use different sources to gather relevant data for their analysis. It is also one of the few studies in political science where a corpus was created to analyze the data.

4.3 Sentiment Analysis

Sentiment analysis is a growing research method at the intersection of linguistics and computer-based automated approaches which attempts to automatically determine the sentiment contained in a certain text (Taboada, 2016). Automated sentiment analysis presents an innovative approach in social sciences, the main aim of which is to measure the polarity or tonality of texts by identifying and assessing expressions that people use to evaluate persons, events, or identities (Haselmayer and Jenny, 2017). Although it is becoming increasingly popular in political science mainly because the digitization of legislative transcripts has increased the potential application of established tools for analyses of emotion in text (Cochrane et al., 2021), many political scientists are still more comfortable using human-based content analysis to analyze emotions. The potential problem of analyzing sentiment in parliamentary debates is that unlike text, speeches consist of intonation, facial expressions and body language which are hard to determine just by looking at the transcripts. Hence coders frequently focus not only on reading the transcripts but also on watching video clips of the debates to grasp emotions in their entirety.

4.3.1 Gender Influence on Negativity in Parliament

Research problem: Haselmayer, Dingler and Jenny (2022) analyzed how the gender of the MPs and the context of debates influenced the level of negativity in parliamentary speeches and showed that female MPs used less negative language than male MPs mainly because of gender differences in socialization and stereotypical expectations.

Data collection: The authors focused their analysis on 52.132 speeches from plenary debates in the Austrian National Council. Those speeches were delivered by more than 500 different MPs from 7 Austrian parties (SPÖ, ÖVP, FPÖ, BZÖ, LiF, Greens and Team Stronach) throughout 24 years (from 1993 to 2013). Speeches from cabinet members (approximately 4.000) and short speeches with less than five sentences (around 500) were excluded from the analysis.

Research method: The authors applied sentiment analysis with word embeddings to plenary speeches in Austrian parliament. They researched negative

³ <https://lumivero.com/products/nvivo/>

parliamentary speeches and relied their analysis on machine learning based on crowd-coded training set. The classifier used data and word embeddings from FastText library⁴ (META, 2015). The authors calculated meaningful word vectors by using subwords and the Gensim library⁵ (LGPL, 2009). Each sentence was represented as a sequence of word vectors which preserved information on word order and captured dependencies between words. They also used a recurrent neural network (The Gated Recurrent Unit – GRU) to deal with a sequential data input. In the stage of pre-processing the text, stop words and punctuation were included. They trained this procedure on around 20,000 sentences which contained a continuous negativity score ranging from neutral to very negative (0 – 4). The model was then trained 60 times with a dropout of 40 % over the entire network.

Discussion: Although this is a political science research, the data collection and analysis descriptions are highly computational and therefore require some computational knowledge to be fully understandable. Since one of the common goals is to familiarize other political and social scientists with automated sentiment analysis, a more simplified description of the methods would be useful.

4.3.2 Emotions in Political Speech

Research problem: Cochrane et al. (2021) analyzed a new dataset of annotated texts and videos from the Canadian House of Commons to examine whether transcripts capture the emotional content of speeches, to compare strategies for the automated sentiment analysis in text and test the robustness of the approach based on word embeddings.

Data collection: Their data collection consisted of official Hansard transcripts and video clips. To gather the latter, the authors recorded every third Question Period in the Canadian parliament between January 2015 and December 2017. This covered the last 10 months of Stephen Harper's conservative and the first 23 months of Justin Trudeau's liberal government. They trimmed the videos from the start of the first question to the end of the last answer which produced 102 videos of approximately 45 minutes in length and randomly selected ten time-points (*mm:ss*) in each of them. The sentence beginning just prior to the time-point was extracted as its own video clip. The average length of the extracted clip was approximately 9 seconds and it contained 23 words. These video clips were added to a Qualtrics⁶ survey instrument and randomly assigned to one of three independent, bilingual coders for manual coding. For all but one video clip the authors were also able to identify the corresponding official Hansard transcriptions. For speeches in French, the coders used the official English translations. The coders were asked to assign a sentiment score to each clip depending on eleven-point scale (0 – 10, negative – positive) as well as activation (subdued – aroused) of the speech

fragment. Since the presentation of clips was randomized, same clips were often presented to the same coder at different times. The texts of the speech fragments were also randomly presented to three independent coders who were asked to indicate the sentiment and activation for each fragment on eleven-point scale. Throughout their analysis, the authors also tested five widely used sentiment dictionaries (Lexicoder 3.0, Sentiwordnet 3.0, Hu-Liu Lexicon, VADER, and Jockers-Rinker's Lexicon) to test their efficacy.

Research method: Researchers employed sentiment analysis with the help of automatically generated sentiment dictionaries. In addition, sentiment was manually coded by coders to improve reliability of the research results.

Discussion: This paper shows that when conducting sentiment analysis, political scientists can rely on video clips of the parliamentary debates and use them to triangulate data gathered from the analysis of official parliamentary transcriptions which improves the reliability of the research.

4.4 Mixed Methods Approach

Mixed methods approach draws on the strengths of qualitative and quantitative research methods which generates a more complete picture of the research problem (Shorten and Smith, 2017). It is a highly complementary approach where the results of one research method can be validated, elaborated, and clarified by the other. Such triangulation allows not only for more valid research results but also reduces research bias and unwarranted selectivity of source materials, which according to Thies (2002) are the two biggest problems of qualitative research. Mixed methods offer more in-depth findings and forces researchers to develop a broader set of research skills which produce valid research results (Tzagkarakis and Kritas, 2022).

Corpus-assisted discourse studies (CADS) could be understood as a special type of mixed methods approach as they combine qualitative discourse analysis with predominantly quantitative corpus-assisted research approach. Rubtcova et al. (2017) show that it is a useful research method for the study of political discourse and parliamentary data especially when the data has already been collected in a corpus (as in ParlaMint). This approach uses corpus techniques to examine a particular political discourse type and analyze certain patterns of language with one of the greatest strengths being minimization and reduction of the research bias (Partington, 2012).

4.4.1 Performance, Gender, and Affective Atmosphere in the time of Brexit

Research problem: Parry and Johnson (2021) examined the parliamentary discourse regarding threats to Members of the Parliament in the context of

⁴ <https://fasttext.cc>

⁵ <https://radimrehurek.com/gensim/>

⁶ <https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/survey-tools-overview/>

broader discussions about emotionality, polarization, and toxicity in discourse in the UK.

Data collection: The primary source material were Hansard transcripts of the debate on September 25. The debate started with the PM's address at 6.30 in the afternoon and ended 3 hours later. They also used data provided by the UK Parliament's YouTube channel; this allowed them to watch relevant sections and capture gestures, use of space and affective atmosphere. In addition, they used the Nexis database⁷ for the newspaper analysis. Here they searched for "Tracy Babin" and "Paula Sherriff" since the names of the two MPs were determined to provide the most relevant results regarding the research topic. They read the articles and retained those that focused on the abuse of the female MPs or those which called for the new standards in public life and language. This news sample comprised 97 articles, mostly from national news outlets.

Research method: The authors employed a mixed methods approach combining performance analysis of the Hansard transcripts and UK Parliament YouTube coverage of the debates and discourse analysis of national as well as local newspaper coverage of the parliamentary debates. Using the performance approach allows the researchers to conduct research beyond the linguistic content of political speech and to focus on style, form, gesture, and the use of physical space.

Discussion: This is not a typical use of the mixed methods approach since the authors did not combine quantitative and qualitative but rather two qualitative approaches. This research is significant also because it is the only one in our sample which employed performance analysis. This paper also shows how important it is to not only focus on one data source but rather combine various sources and different types of data.

4.4.2 Religious Freedom in Debates on Same-sex Marriage in Australia

Research problem: Poulos (2019) explored why and how the term "religious freedom" appeared in the title of the Australian bill to legalize same-sex marriage. He wanted to analyze how debates about same-sex marriages changed over time.

Data collection: Poulos analyzed 663 speeches made in Australian parliament during the marriage legislation debates between 2004 and 2017. This research was based on 20 bills proposing amendments to the Marriage Act allowing for same-sex marriage or recognizing same-sex marriages. Data was taken from the Australian Parliament House website using the homepages of the respective bills as well as the Hansard. Once the same-sex marriage bills were identified, PDFs of the Hansard files were collected for every speech and then converted to the text file using an online converter.⁸ Poulos removed

all the metadata (speakers' names, electorates, ministerial roles, time stamps, etc.), interjections and procedural statements included in the Hansard files. The speeches were chronologically grouped into three different sub-corpora (the first one from 2004, the second one between 2006 and 2016 and the third one from 2017). Then, two other sub-corpora were created, this time according to whether the speakers explicated a position in support or in opposition to the same-sex marriage and then chronologically sorted again according to support or the opposition.

Research method: The author opted for corpus-assisted discourse analysis. Poulos analyzed the text files with the help of two software packages, namely AntConc⁹ (Anthony, 2018) and WordSmith Tools¹⁰ (Oxford University Press, 1996). The first was used to generate word frequency lists, concordances, and identify collocates and the second one to identify keywords. This analysis was triangulated with manual discourse coding using the NVivo software. To examine whether the arguments were framed for or against the same-sex marriage, each sub-corpus of the supportive speeches was analyzed against corpus, which included the speeches which opposed same sex marriage and vice versa. The author examined the most frequent words and lexical keywords from each of the sub-corpora and performed the analysis of how the framing of the same-sex marriage "issue" changed over time.

Discussion: This is a rare example of research which deliberately discarded the available metadata. This is uncommon in social sciences which usually relies on metadata to provide additional information during analysis.

4.5 Thematic Analysis

Thematic analysis is a highly useful approach in qualitative research since it allows for the identification of prominent themes and provides several ways to interpret meaning from a certain dataset. Its focus is to find not only the major themes of analyzed data, but also to come up with various fine-grained subthemes that match the main themes and therefore make the interpretation of results much more straightforward (Gherghina, Tap and Soare, 2022). It is commonly understood as an umbrella term for various research approaches rather than a single method. In political studies, thematic analysis (sometimes referred to as qualitative document/content analysis) is particularly useful for the study of legislation and policy and is also becoming increasingly important in the study of parliamentary debates.

Sometimes thematic analysis is equated to content analysis and much of this confusion is because thematic analysis originated from content analysis before branching off to serve similar but distinct research goals (Joffe, 2012). The main difference between the two lies in the possibility of quantification

⁷<https://www.lexisnexis.com/en-us/professional/nexis/nexis.page>

⁸<https://pdftotext.com>

⁹<https://www.laurenceanthony.net/software/antconc/>

¹⁰<https://www.lexically.net/wordsmith/>

of data in content analysis by measuring the frequency of categories and themes, whereas thematic analysis is strictly qualitative. Consequently, content analysis has a wider selection of coding approaches, is more practical and straightforward whereas thematic analysis supports deeper immersion and is more intuitive.

4.5.1 Parliamentary Debates About Emigrants

Research problem: Gherghina, Tap and Soare (2022) analyzed the ways in which members of the Romanian parliament refer to emigrants; not only the ambivalent attitude but also the representation of emigrants and their needs.

Data collection: The authors focused on analyzing parliamentary speeches from the plenary sessions in the Chamber of Deputies (lower house of the Romanian parliament) in the two terms between 2012 – 2016 and 2016 – 2020 with an incomplete second term (data was available only until March 2020 whereas the term ended in November 2020). This yielded 239 parliamentary speeches which covered the developments after the financial crisis and important events (elections, anti-government protests) in which the diaspora actively participated. The speeches were split between the two terms as follows: 135 speeches with the average length of 530 words from the first term and 104 speeches with average length of 517 words from the second term. The speeches were publicly available on the official website of the Chamber of Deputies. Before the analysis, data was coded in three stages. First, coders independently read all relevant speeches and grouped them into predefined themes. Second, an inter-coder reliability test was used to identify borderline and missing themes. In the final phase, the list of main themes was enriched with the relevant sub-themes and applied to the speeches.

Research method: The authors employed deductive thematic analysis based on the pre-established themes which were derived from the literature. This allowed for the identification of comment themes as well as provided various ways to interpret meaning from the dataset of speeches selected for the analysis.

Discussion: One shortcoming that authors mention is an underrepresentation of Romanian emigrants in Romanian politics which could influence the content of speeches about the diaspora. In addition, not all the speeches were collected which could have some impact on reliability of the research results.

4.6 Text as Data and Computational Approaches

Computational methods have in the last couple of years gained in popularity which allowed for the development of new research approaches and new methods to analyze textual documents inside social sciences. One such is **text-as-data approach** which consists of a broad set of techniques and relies on automated or semi-automated analysis of text (Gilardi

and Wüest, 2020). It allows researchers to analyze extensive amounts of textual data, significantly reduces the cost of analyzing large collections of text and allows researchers to deploy language-agnostic analytical tools. Text-as-data is a relatively new approach in political science in comparison to the more traditionally used content analysis and qualitative methods (Krippendorff, 2018). It combines new sources of data, machine-learning tools, and social science research design to develop and evaluate new insights (Grimmer, Roberts and Steward, 2022) and understands text as numerical data suitable for quantitative analysis. The aim of this approach is not to replace the insights of qualitative research but rather complement and extend it (Mochtak, personal communication, 2023).

Quantitative text analysis (QTA) is an example of the text-as-data approach and refers to the process of analyzing text data by using statistical procedures. It is an automated and systematic method for processing extensive amounts of text (e.g., parliamentary debates, policy documents, party manifestos, etc.) (Slapin, 2018) which most commonly occurs in three basic steps: 1) defining a corpus from the texts for analysis, 2) determining the unit of analysis, and 3) creating document feature matrix.

Social network analysis (network analysis) refers to the study of social structures by using networks and graph theory. It analyzes links between nodes, which in political science most commonly represent either persons, organizations, or states while links represent some form of connection between them (Ward, Stovel and Sacks, 2011). Social network analysis is becoming an increasingly used computational method in political science and is commonly used when researchers want to establish connections between political actors from an extensive dataset. As shown in Skubic et al. (2022), network analysis can be extremely useful for the comparative analysis of argumentative and structural power of parliamentarians in different European parliaments.

4.6.1 Populism and Parliamentary Polarization in German Parliament

Research problem: Lewandowski et al. (2021) examined how the German parliamentary discourse changed after two populist and two non-populist parties entered parliament and analyzed how populism shaped the behavior of new parties as well as how other parties respond when the new contesters arrive.

Data collection: The authors based their analysis on a GermaParl corpus¹¹ (Blätte and Blessing, 2018) which includes parliamentary debates from the German parliament. They analyzed legislative periods 9 to 19 (from 1980 to 2020) and focused on two populist (The Left, AfD) and two non-populist parties (Greens, PDS). Only speeches delivered by members of the parliament were analyzed and speakers not belonging to a parliamentary group were

¹¹ <https://github.com/PolMine/GermaParlTEI>

excluded. The analysis of populist language was based on all speeches from the period of interest (190.000) whereas the analysis of polarization was based on a subset of approximately 113.000 speeches. When measuring parliamentary polarization, the authors only included those speeches to which they could assign a substantial topic using a topic modelling approach which resulted in a lower number of analyzed speeches.

Research method: They applied qualitative text analysis of parliamentary speeches to measure populism and issue-based polarization. To measure populist speech, they used a dictionary-based approach. Firstly, they used a specific word list (suggested by Rooduijn and Pauwels, 2011) to create a lexicon of key terms which indicated the use of populist references. Then they calculated the frequency of those terms relative to the length of a speech as well as used keyword-in-context analysis to examine the context in which the identified words occurred. For measuring political polarization, the authors used the Wordfish algorithm¹² (Slapin and Proksch, 2008) for which they needed to subset all speeches along three dimensions: the parliamentary group, the primary topic of the speech and the legislative period in which the speech was made. All speeches of a single parliamentary group in each legislative period were clustered about a single topic.

Discussion: This is the only reviewed research which uses Wordfish algorithm, which is written in the programming language for statistical computing R. As shown later in the discussion chapter, R is especially important for political science since it is easy to understand and provides data in tabular format and is therefore most used programming language for extracting political positions from textual documents.

4.6.2 Analyzing the Politics of Brexit Debate Abroad

Research problem: Sierens and Brack (2020) examined to what extent the attention given to Brexit differs across different parliaments and if parties emphasized the same issues across different levels. They specifically analyzed how Brexit was framed and discussed in the Belgium parliament.

Data collection: Research relied on a unique database of parliamentary questions in three different Belgian parliamentary assemblies (Federal, Flemish, and Walloon). The authors gathered data from January 23, 2013 (when David Cameron announced his intention to hold a referendum about Brexit) until October 2017. Data for analysis were retrieved directly from the websites of all three assemblies. At the federal level, the authors analyzed parliamentary questions asked in the Chamber of Representatives and used the keyword “Brexit” to classify all questions that dealt with this specific topic. At the regional levels, they focused on questions that had the word “Brexit” in their titles. Altogether, they retrieved 146 parliamentary questions in the Federal parliament (94 oral and 52 written), 88 parliamentary questions in

Flanders (57 oral and 31 written) and 37 parliamentary questions in Wallonia (12 oral and 25 written). For the purpose of comparative analysis, the authors categorized data into series of questions divided into “who” questions (“who asks who?”, “who asks what?”, etc.) and “what” questions. The former were classified according to the MPs party and presence/absence in the governmental coalition. In the latter, each parliamentary question was categorized according to the main issue emphasized in the parliamentary question. According to these criteria the data was coded into four most frequent categories (general information on Brexit, trade and economic consequences, negotiation strategy, specific issues).

Research method: In the first step, the authors conducted a descriptive comparative analysis of the gathered parliamentary questions. It relied on Social Network Analysis that allowed the authors to focus on the structural relationships between the different units of analysis. For each level of the government, they drew networks of parliamentary questions and computed various indicators of those networks (density, average degree, etc.). In the second step they employed loglinear modelling (a special case of generalized linear models for multivariate cross-classified categorical data (Sierens and Brack, 2020)) of the frequency of associations and interactions between categorical variables.

Discussion: This is the only reviewed paper that employs social network and loglinear model analysis. Although authors provide some explanation of the methods, there is no emphasis on a more detailed explanation (e.g., which software was used for network analysis, how to work with such software, etc.).

5. Discussion and Conclusion

In this literature review we showed the most common methods and approaches political scientists use when conducting research on parliamentary debates and discourse. One of the core interests of political science is to analyze power relations inside parliaments as well as a means through which the power is displayed. Parliamentary discourse not only reflects the power and authority of the parliamentarians, but also allows parliamentarians to present their interpretation of specific issues to different external audiences (Laver et al., 2003).

One of our main findings is the similarity between methods and approaches used in political science and sociology, as shown in Skubic and Fišer (2022). Our extensive research showed that more than half of the reviewed political science papers employed one of the research methods that are traditional in political science (either content or discourse analysis). We also find that political scientists often employ such methods to analyze data which is frequently manually collected and downloaded from various sources (e.g., parliamentary websites, repositories of relevant

¹² <http://www.wordfish.org>

organizations, etc.) rather than using more modern and less time-consuming and resource heavy computational techniques. This is confirmed by Mochtak (personal communication, 2023) who states that more than 90 % of political science research still employs traditional data collection and research methods (with content analysis being the most used). According to Mochtak, political science is slow when it comes to adjustment and modification of research methods and approaches to more modern, less time-consuming, and more technologically advanced methods. Our review shows that in some cases this transition has already been made but such research is scarce, hard to identify, and often lacks methodological explanation.

Even when political scientists use modern computational methods to conduct research and collect data, they are often reluctant to perform big-corpora and big-data analyses or employ methods which they find hard to comprehend. According to Mochtak, political scientists only rarely rely on complicated programming language or computational methods. Probably one of the most used in political science is R programming language mainly because it offers tidy data in tabular format and is relatively easy to use. When political scientists deal with large amounts of quantitative data, they want them to be organized, easily accessible and easy to use (one example of such data is V-DEM data¹³). Databases therefore need to be made approachable, accessible and offer functional API for political scientists to consider using them.

We find that despite the quantitative turn of political science in recent years, political scientists still predominantly use qualitative or mixed methods. In addition, software and tools for computational qualitative analyses (such as Nvivo, Atlas.ti or MAXQDA) have in recent years become more popular. This not only allows researchers to analyze data faster, more efficiently and in a more organized way but also attributes to more replicable and relevant research results and minimizes researcher bias which is otherwise common in solely qualitative research. Reliability and relevance of results is further enforced by data collection triangulation which is common in political science. Often researchers rely not only on parliamentary but also other sources such as newspapers, (social) media, interviews, etc., which assures higher quality of the conducted research.

If we want to encourage political scientists to start incorporating corpora such as ParlaMint in their research and use corpus-assisted methods more actively, we firstly need to make it highly approachable and accessible (Mochtak, personal communication, 2023). Datasets such as ParlaMint are very useful and offer an abundance of valuable data but are often too complex for political scientists to use. Our first aim should therefore be to make data available in a format which political scientists would be familiar with. In addition, tutorials, workshops, showcases, and user manuals should be offered to

political scientists so they could familiarize themselves with the ParlaMint concept, workflow, and the variety of data it offers. We agree with Kytö (2011) that corpus compilers should also provide rich, useful, and user-friendly documentation as to how the corpus data is gathered, processed, and annotated and should clearly and in detail document their compilation decisions, offering user guides, corpus manuals and training materials which would accompany the release versions of corpora. This would enable political scientists to reuse corpora in a contextualized way, which would significantly ease their process of data collection and analysis.

In addition, the ParlaMint community should also focus on providing data with rich and useful metadata. Metadata such as gender, role, party affiliation, political orientation, etc. are useful, but other metadata such as sentiment score, emotions, policy areas of agenda points etc. would be an additional added value. Collecting and assigning such metadata is usually a time-consuming process which requires a lot of effort and human resource and is frequently very specific to the research question at hand. This is why corpora such as ParlaMint would be even more interesting for political scientists if it allowed them to directly add, edit and share additional metadata layers. Machine translations of parliamentary debates would also provide important additional possibilities for more international research and parliamentary discourse comparisons.

The argument that we want to put forward with this literature review is not that the current predominantly qualitative research methods in political science should be replaced with more quantitative corpus-assisted approaches in their entirety, but rather that corpus data and corpus-analytical techniques could effectively be used alongside the traditional qualitative approaches. We understand corpora as potentially powerful tools which would help political scientists not only to simplify data collection processes and help them generate relevant results much more effortlessly but would also contribute to more transparent, verifiable, and reproducible research.

6. Acknowledgements

The work described in this paper was funded by the Slovenian Research Agency research programme P6-0436: Digital Humanities: resources, tools, and methods (2022-202), the CLARIN ERIC ParlaMint project (<https://www.clarin.eu/parlamint>) and the DARIAH-SI research infrastructure (<https://www.dariah.eu>).

7. Bibliographical References

Anthony, L. (2018). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>

¹³ <https://github.com/vdeminstitute/vdemdata>

- Blassnig, S. (2022). Content Analysis in the Research Field of Political Communication: The Self-Presentation of Political Actors. In *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft—Standardized Content Analysis in Communication Research: Ein Handbuch-A Handbook* (pp. 301-312). Wiesbaden: Springer Fachmedien Wiesbaden.
- Blätte, A., & Blessing, A. (2018, May). The GermaParl corpus of parliamentary protocols. In *proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- Blaxill, L. (2022, June). Parliamentary Corpora and Research in Political Science and Political History. In *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference* (pp. 33-34).
- Carroll, S. J. (1984). Woman candidates and support for feminist concerns: The closet feminist syndrome. *Western Political Quarterly*, 37(2), 307-323.
- Cochrane, C., Rheault, L., Godbout, J. F., Whyte, T., Wong, M. W. C., & Borwein, S. (2022). The automatic analysis of emotion in political speech based on transcripts. *Political Communication*, 39(1), 98-121.
- Dunmire, P. L. (2012). Political discourse analysis: Exploring the language of politics and the politics of language. *Language and Linguistics Compass*, 6(11), 735-751.
- Erjavec, T., Ogrodniczuk, M., Osenova, P., Ljubešić, N., Simov, K., Pančur, A., ... & Fišer, D. (2023). The ParlaMint corpora of parliamentary proceedings. *Language resources and evaluation*, 57(1), 415-448.
- Farrelly, M. (2010). Critical discourse analysis in political studies: An illustrative analysis of the 'empowerment' agenda. *Politics*, 30(2), 98-104.
- Cauchon, S. Introduction to Political Science Research Methods. eBook – Adobe PDF. 1st Edition. <https://ipsrm.com/>.
- Gherghina, S., Tap, P., & Soare, S. (2022). More than voters: Parliamentary debates about emigrants in a new democracy. *Ethnicities*, 22(3), 487-506.
- Gilardi, F., & Wüest, B. (2020). Using text-as-data methods in comparative policy analysis. In *Handbook of research methods and applications in comparative policy analysis* (pp. 203-217). Edward Elgar Publishing.
- Haselmayer, M., & Jenny, M. (2017). Sentiment analysis of political communication: Combining a dictionary approach with crowdcoding. *Quality & quantity*, 51, 2623-2646.
- 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.
- Jarvis, L. (2022). Constructing the coronavirus crisis: Narratives of time in British political discourse on COVID-19. *British Politics*, 17(1), 24-43.
- Joffe, H. (2012). Thematic analysis. *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners*, 209-223.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Kuckartz, U. (2019). Qualitative text analysis: A systematic approach. *Compendium for early career researchers in mathematics education*, 181-197.
- Kytö, M. (2011). Corpora and historical linguistics. *Revista Brasileira de Linguística Aplicada*, 11, 417-457.
- Lauer, J. (2021). Methodology and political science: the discipline needs three fundamentally different methodological traditions. *SN Social Sciences*, 1(1), 43.
- 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.
- Lewandowsky, M., Schwanholz, J., Leonhardt, C., & Blätte, A. (2022). New parties, populism, and parliamentary polarization: evidence from plenary debates in the German Bundestag. *The Palgrave handbook of populism*, 611-627.
- Lilja, M. (2021). Russian political discourse on illegal drugs: A thematic analysis of parliamentary debates. *Substance Use & Misuse*, 56(7), 1010-1017.
- May, P. (2016). Ideological justifications for restrictive immigration policies: An analysis of parliamentary discourses on immigration in France and Canada (2006–2013). *French Politics*, 14, 287-310.
- Och, M. (2019). Conservative feminists? An exploration of feminist arguments in parliamentary debates of the bundestag1. *Parliamentary Affairs*, 72(2), 353-378.
- Parry, K., & Johnson, B. (2023). Humbug and outrage: A study of performance, gender and affective atmosphere in the mediation of a critical parliamentary moment. *The British Journal of Politics and International Relations*, 25(1), 3-20.
- Partington, A. "Corpus analysis of political language." In: C.A. Chapelle (Ed.) *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd. (2013).
- Poulos, E. (2020). The power of belief: religious freedom in Australian parliamentary debates on same-sex marriage. *Australian Journal of Political Science*, 55(1), 1-19.

- Provalis Research. (2014). QDA Miner.
- Sanchez Salgado, R. (2021). Emotions in European parliamentary debates: Passionate speakers or unemotional gentlemen?. *Comparative European Politics*, 19(4), 509-533.
- Scott, Mike. "WordSmith Tools Manual." (2018): <https://lexically.net/downloads/version7/HTML/>
- Shorten, A., & Smith, J. (2017). Mixed methods research: expanding the evidence base. *Evidence-based nursing*, 20(3), 74-75.
- Sierens, V., & Brack, N. (2021). The politics of the Brexit debate abroad: an analysis of parliamentary questions on Brexit in Belgian parliaments. *Journal of Contemporary European Studies*, 29(4), 519-534.
- Skubic, J., & Fišer, D. (2022, June). Parliamentary discourse research in sociology: Literature review. In *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference* (pp. 81-91).
- Skubic, J., & Fišer, D. (2022). "Parliamentary Discourse Research in History: Literature Review." In *Proceedings of the Conference on Language Technologies and Digital Humanities*.
- Skubic, J., Angermeier, J., Bruncrona, A., Evkoski, B., & Leiminger, L. (2022). Networks of power: Gender analysis in selected european parliaments. In 2nd Workshop on Computational Linguistics for Political Text Analysis (CPSS).
- Slapin, J. (2020). Three Basic Steps of Quantitative Text Analysis.
- 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.
- Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, 2, 325-347.
- Thies, C. G. (2002). A pragmatic guide to qualitative historical analysis in the study of international relations. *International Studies Perspectives*, 3(4), 351-372.
- Tzagkarakis, S. I., & Kritas, D. (2022). Mixed research methods in political science and governance: approaches and applications. *Quality & Quantity*, 1-15.
- Van Dijk, T. A. (2018). Discourse and migration. *Qualitative research in European migration studies*, 227-245.
- Van Dijk, T. A. (1997). What is political discourse analysis. *Belgian journal of linguistics*, 11(1), 11-52.
- Ward, M. D., Stovel, K., & Sacks, A. (2011). Network analysis and political science. *Annual Review of Political Science*, 14, 245-264.