Bachelor's Thesis

On the difficulty of using Contextual Word Embeddings to Measure Political Polarisation in Parliamentary Speech

Author
Giulio Cusenza
giulio.cusenza@student.uni-tuebingen.de
giuliocusenza@gmail.com

Supervisor
Çağrı Çöltekin
cagri.coeltekin@uni-tuebingen.de

A thesis submitted in partial fulfilment
of the requirements for the degree of

Bachelor of Arts
in
International Studies in Computational Linguistics

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen

October 2024

Antiplagiatserklärung

Ich erkläre hiermit,

dass ich die vorliegende Arbeit selbständig verfasst habe,

dass ich keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe,

dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe,

dass die Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist,

dass ich die Arbeit weder vollständig noch in wesentlichen Teilen bereits veröffentlicht habe,

dass das in Dateiform eingereichte Exemplar mit dem eingereichten gebundenen Exemplar übereinstimmt.

Anti-Plagiarism Statement

I hereby declare

that I have written this work independently,

that I have not used any resources or sources other than those specified,

that I have identified all statements taken verbatim or in essence from other works as such,

that the work has not been the subject of any other examination procedure, either in full or in substantial parts,

that I have not already published the work, either in full or in substantial parts,

that the copy submitted in file form corresponds to the bound copy submitted.

Abstract

This study attempted to measure political polarisation in parliamentary speech using contextual word embeddings, replicating a method introduced by Ding et al. (2023). Data from the UK's House of Commons and the Italian Senate were analysed, focusing on selected topical words across political parties. Contrary to the original findings, the technique failed to reveal meaningful patterns of polarisation between parties in either dataset. Various factors are highlighted to explain the failure, including the unique properties of parliamentary speech compared to broadcast media speech, such as its debate structure and formal register. Technical complications with the data and language models are also discussed. The study concludes that contextual embeddings may not be suitable for this type of analysis due to their black-box nature and anisotropic behavior.

Contents

1 Introduction **1**

2 Data **2**

3 Method **5**

4 Results **7**

5 Discussion **10**

6 Conclusion **11**

7 Materials **11**

1 Introduction

Since the field of Natural Language Processing (NLP) was started, there has been significant interest in employing its techniques in various political settings, whether for the analysis of political speech or as tools for politicians themselves. More recently, the exacerbation of public debate has raised the questions of a definition and a measure of political polarisation (Lelkes, 2016), with NLP researchers exploring several different methods (Németh, 2023). A particularly intriguing approach was proposed by Ding et al. (2023). Their strategy revolves around using the language model BERT (Devlin et al., 2019) to obtain contextual word embeddings for topical words from two major broadcasting companies with opposing political alignments (CNN and FoxNews), the hypothesis being that the same topical words (e.g., *racism*, *immigration*, *police*, etc.) would be employed differently by the two broadcast networks. This difference can thus be quantified through the cosine similarity between the embeddings, providing a measure of polarisation.

A review by Németh (2023) finds that only 20% of NLP studies on political polarisation use congressional or parliament speeches as their data sources. While broadcast and social media are undeniably important to understand social phenomena, it is likewise important to investigate the extent and influence of polarisation in key decision-making institutions. This study attempts to apply the method from Ding et al. (2023)'s study to parliamentary data instead of broadcast media. It is worth noting that, despite the authors' claim to include the study's code framework in their GitHub account, there is no mention of the account's name or URL, nor was it possible to contact the authors through the provided e-mail addresses. Therefore, while the technique may be the same, the details of the implementation are necessarily different, hindering the full reproducibility of the study.

The initial plan was to refine the method on Italian data, and then potentially repeat the analysis across multiple countries. This is not as straightforward as it may seem: while in the United States and other countries there are two traditional opposing parties still dominating the political scene, many other Western countries have been experiencing a severe party fragmentation (Mair, 2023) and electoral volatility (Blumenstiel and Plischke, 2015), meaning that there are more than two major parties and the distribution of votes varies substantially at each election. Therefore, identifying the candidate "poles" for such an analysis on polarisation is no trivial task, and requires a thorough understanding of each country's political scene. The high electoral volatility also complicates diachronic analysis especially for a study on parliamentary debates, as the amount of data available for each political group shrinks and stretches by each legislative term based on the number of elected representatives.

As the Italian data presented these as well as some technical problems (see Section 2), the study was also conducted on data from the United Kingdom's House of Commons. These data offer two primary advantages: first, the UK's two traditional political parties remain relatively strong; second, the availability of English-language data allows for the application of more sophisticated – or at least more extensively documented – language models.

All the code produced for this study, as well as the processed data and generated figures, are available in a GitHub repository referenced in Section 7 (“Materials”).

2 Data

ParlaMint[1] is a project by CLARIN[2] providing “comparable corpora of parliamentary debates of 29 European countries and autonomous regions, covering at least the period from 2015 to 2022”. This study utilises the corpora from the Senate of the Italian Republic (from March 2013 to September 2022) and the Parliament of the United Kingdom (from January 2015 to July 2022) included in ParlaMint’s 4.1 release (Erjavec et al., 2024). From the latter, only the speeches from the House of Commons were kept for analysis, as the House of Lords is not democratically elected and as such can not be easily grouped into political parties.

It is important to highlight the particular properties of parliamentary data compared to the broadcast media language used in the study by Ding et al. (2023). One notable distinction is quantitative: parliaments typically hold only a few sessions per week, with part of these sessions devoted to formalities and voting. In contrast, broadcasting companies often operate multiple television or radio channels with uninterrupted streams of news and talks, resulting in a significantly larger volume of data over the same time frame. Table 1 presents the number of speeches and words included in each dataset (including the House of Lords in the case of the UK). The lower numbers in the Italian data are likely due to the lack of data from the larger house of the Italian Parliament, the Chamber of Deputies.

dataset	**n speeches**	**n words**
ParlaMintGB	670,912	126,705,494
ParlaMintIT	172,796	31,970,163

Table 1: Number of speeches and words of each dataset.

Tables 2 and 3 show the total number of occurrences of the topical words selected for this study over the available time frames, as well as the number of sessions in which each word was pronounced and the word’s per-session frequency, which indicates how concentrated the debate around a given topic is. Figures 1 and 2 illustrate the occurrences’ yearly break-down. Note that the counts for the first and last year of each graph may not be complete, as the datasets are built spanning over precise legislative

[1] `clarin.eu/parlamint`
[2] `clarin.eu`

terms (e.g., the Italian dataset starts in March 15th 2013, with the beginning of the XVII legislature). Some of the graphs reveal interesting similarities. For instance, in both cases, the topic of immigration peaks in 2018, while mentions of energy and Russia increase sharply following Russia's full-scale invasion of Ukraine in 2022.

topic	**occurrences**	**sessions**	**occurrences per session**
education	21,262	1,062	20.0
energy	13,679	1,004	13.6
Europe	11,440	1,079	10.6
immigration	6,495	764	8.5
Russia	5,536	651	8.5
unemployment	3,545	737	4.8

Table 2: Number of times a specific word was pronounced in the UK's House of Commons from March 2013 to September 2022, including the number of sessions in which that word was pronounced and the word's per-session frequency.

topic	**occurrences**	**sessions**	**occurrences per session**
scuola	4,532	651	7.0
energia	2,221	452	5.0
Europa	10,969	909	12.1
immigrazione	1,552	335	4.6
Russia	903	226	4.0
disoccupazione	1,025	328	3.1

Table 3: Number of times a specific word was pronounced in the Italian Senate from from January 2015 to July 2022, including the number of sessions in which that word was pronounced and the word's per-session frequency.

Returning to the comparison with broadcast media, it is the qualitative differences that are most significant: language used in broadcast news reports differs substantially from that of elected politicians discussing major issues. Crucially, in the case of politically aligned broadcasting companies we can observe the phenomenon of *agenda setting* (McCombs and Shaw, 1972): the amount of news coverage an event receives determines how much importance the public audience attaches to that event. This can also differ for politically opposing broadcasters; for example, the coverage of news concerning immigration could be more frequent on a specific side of the political spectrum. The framing of news (Goffman, 1974; Nelson et al., 1997) is also a key aspect: right-wing broadcasters may discuss immigration in the context of a crime committed by an immigrant, while left-wing outlets may focus more on human rights in detention centres. This dynamic is absent or much more limited in parliamentary debates, where all parties contribute to the selection of the issues to be discussed and participate in the same debate. This debate structure greatly limits the variation of context in the different groups.

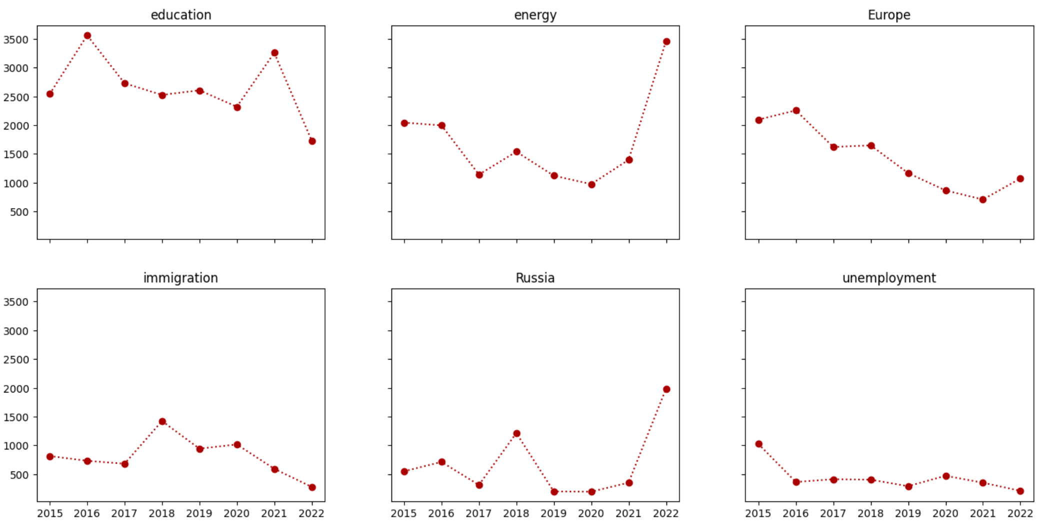

Figure 1: Number of times selected topical words were pronounced by major parties in the UK's House of Commons from 2015 to 2022.

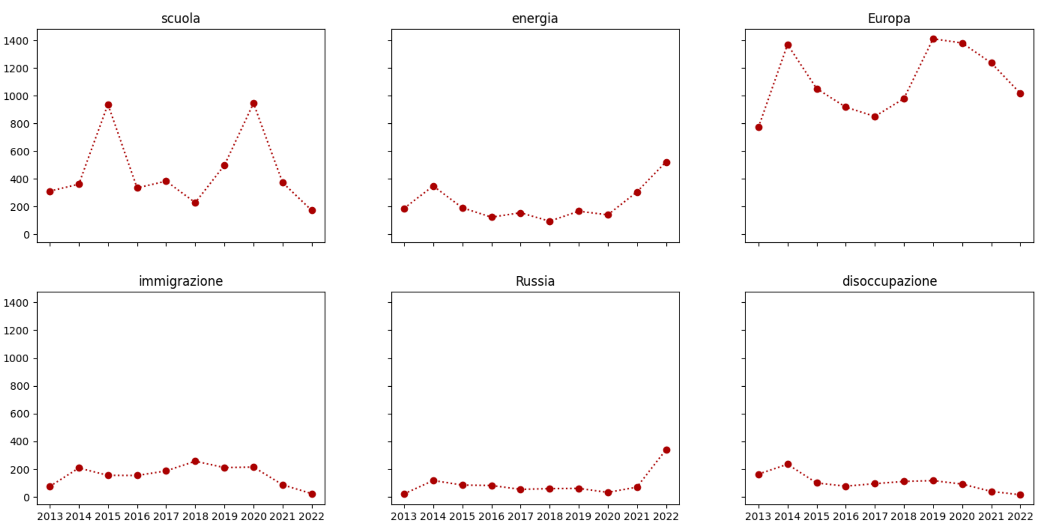

Figure 2: Number of times selected topical words were pronounced by major parties in the Italian Senate from 2013 to 2022.

Another qualitative difference lies in the register: parliamentary speech is often characterized by sophisticated formal rituals and grandiloquent expressions, which are definitely a minority in the training data of most general language models. Critique, mockery, sarcasm, and other forms of reporting the adversary's speech can also pose challenges for contextual embedding models, as these elements introduce words to the context that do not necessarily belong to the current speaker's opinion.

It is important to note that the Italian dataset presented significant challenges due to pervasive structural flaws in its annotation and formatting. These issues primarily arose from difficulties in parsing the plain text files written by official stenographers and provided by Italian institutions. Specifically, the parser struggled to correctly distinguish individual utterances: in some cases, entire dialogues involving multiple speakers were incorrectly grouped into a single segment, including the names and affiliations of the speakers within the text. Ideally, these utterances should instead be separated into different segments, with the names of the speakers stored as XML attributes. This is a critical issue because the other speakers' utterances – and even their names – can affect the contextual embeddings (e.g., surnames like that of Senator Alessandra Mussolini can considerably influence the embedding model for obvious reasons). This issue could not be addressed; however, the same analyses were conducted on the House of Commons data as well.

Another issue had to do with guest speakers. Speakers are categorized as either `#guest` or `#regular`, with the latter tag assigned to elected representatives of the Senate. However, many guest speakers – typically government officials – are marked as `#guest` only at their first intervention in a session, while their subsequent utterances are incorrectly labeled as `#regular`. This issue was easily resolved through a simple script provided together with the materials of this study (Section 7).

3 Method

As mentioned earlier, this study aims to replicate the results of Ding et al. (2023) on parliamentary speech. While their study focused on two selected groups – Democrats and Republicans – through US broadcasting companies CNN and Fox News, most countries do not follow a two-party system, meaning that the analysis of each political landscape calls for country-specific arrangements. In this study, the two traditional parties – Conservative (CON) and Labour (LAB) – were selected for the House of Commons data. In the case of the Italian Senate, five major parties were chosen for analysis: Forza Italia (FI), Fratelli d'Italia (FdI), Lega (LN), Movimento 5 Stelle (M5S), Partito Democratico (PD). Sentences uttered by MPs from other political parties were simply ignored.

In Italy, elected MPs form parliamentary groups based on their coalitions and shared political views. Typically, this leads to major parties attracting "satellite" MPs from much smaller parties. These groups change each term, though a continuity can be traced through the main party within each group. Therefore, in this study, for the sake of simplicity, each parliamentary group was renamed after its dominant party. The party Lega was treated as the successor of Lega Nord.

Next, a set of topical words was arbitrarily chosen, drawing inspiration from Ding et al. (2023)'s original set. For the House of Commons, these are *education*, *energy*, *Europe*, *immigration*, *Russia*, *unemployment*. The same ones, roughly translated into Italian, were chosen for Italian Senate: *scuola*, *energia*, *Europa*, *immigrazione*, *Russia*, *disoccupazione*. However, this approach can be problematic in a low-data setting, as these topics are often referenced using synonyms (e.g., "migration"), periphrases or euphemisms (e.g., "migratory crisis"), or morphologically related words (e.g., "immigrants"). This last issue point is particularly challenging in morphologically rich languages such as Italian. Lemmatisation does not constitute a viable solution, as the embeddings will still be influenced by the different morphosyntactic features, or even semantic differences in cases like *energy* and *energies*), meaning that different forms of the same lemma can not be grouped for analysis. As a result, a sizable portion of the available data is filtered out, including only those utterances containing a word that exactly matches – aside from capitalisation – a topical word in the set.

After extracting the relevant utterances, contextual embeddings were computed for the topical words, with each embedding stored along with the year and the name of the group (i.e., political party of the speaker). To achieve this, the models `bert-base-uncased`[3] and `bert-base-italian-uncased`[4] were used as tokenisers and embedders for the English and Italian data, respectively. The embeddings used in this study are the sum of the last four layers produced after feeding the sentences into the model.

A significant number of words appeared to be missing from the Italian tokeniser's vocabulary, possibly due to the formal register of parliamentary discourse. For instance, the phrase "l'apertura, per quanto cauta - diciamo pure timida - del Consiglio Europeo ad un riorientamento" was split into the following tokens: `l`, `'`, `apertura`, `,`, `per`, `quanto`, `cau`, `##ta`, `-`, `diciamo`, `pure`, `timi`, `##da`, `-`, `del`, `consiglio`, `europeo`, `ad`, `un`, `rio`, `##rie`, `##nta`, `##mento`; not recognising the words meaning *cautious* (feminine), *shy* (feminine), and *reorientation*. However, all topical words are included in the models' dictionaries.

Ding et al. (2023) define a semantic polarity measure SP as follows:

$$SP(A_t, B_t) = 1 - \frac{\sum_{w_i^a \in A_t, w_j^b \in B_t} \cos(w_i^a, w_j^b)}{|A_t||B_t|}$$

where:

t: a specific year (for diachronic analysis)
A_t: set of word embeddings from group A in year t
B_t: set of word embeddings from group B in year t

[3]`huggingface.co/google-bert/bert-base-uncased`
[4]`huggingface.co/dbmdz/bert-base-italian-uncased`

This measure can essentially be read as one minus the average cosine distance between all pairs (w_i^a, w_j^b), where the idea is that the farther the embeddings in group A are from those in group B, the more polarised the two groups are. The score ranges from 0 to 2, which, at least geometrically, correspond to no polarisation and complete polarisation, respectively. However, an $SP > 1$ is only possible with a significant number of negative cosines, which are rare and often difficult to interpret in the case of word embeddings. Since the embeddings being compared here are from the same topical words – with the only difference being that they are used by different political groups – it is extremely unlikely to get even a single negative cosine. Therefore, we will only observe SP values between 0 and 1. While this measure is useful for comparison (e.g., SP across different years), it does not hold any meaning as an absolute value. Further discussion on the limitations of this measure and why it may not be suitable can be found in Section 5.

4 Results

If an effect of the political group were captured by the contextual embeddings, a three-dimensional projection obtained through principal component analysis (PCA) should reveal regions where the embeddings from each political group cluster together. These clusters may be blurred and more or less overlapping, but if the effect is strong, they should still be clearly distinguishable. However, no such clusters were observed in the PCA projections. For instance, Figure 3 shows the PCA results for the embeddings for the word *immigration* in both the English and Italian data for the year 2016. The projections are labeled by political group, but no distinct regions are apparent. Similar results were found across other years and topical words.

Projecting a multidimensional space onto three dimensions inevitably results in information loss. Several tests were conducted using three-component PCA, and in all cases, the PCA never described more than 30% of the variance. The semantic polarity measure, as defined in Section 3, offers a more accurate way to determine whether political group membership affects the distribution. The approach adopted here is to compare the semantic polarity within a group A – $SP(A, A)$ – to the semantic polarity between A and any other group: $SP(A, B)$, $SP(A, C)$, $SP(A, D)$, and so on. If group A uses a word differently than the other groups, $SP(A, A)$ should be lower than the other semantic polarity values. Geometrically, this corresponds to the earlier visual search for clusters in three-dimensional space: we are measuring whether the embeddings in group A are closer to each other than to the embeddings from other groups. The difference now is that we obtain a precise measure, operating in the full multidimensional space, thus avoiding the information loss that comes with dimensionality reduction.

Although mathematically equivalent, it may not be very intuitive to describe intra-party dynamics in terms of polarisation. A more meaningful interpretation of $SP(A, A)$

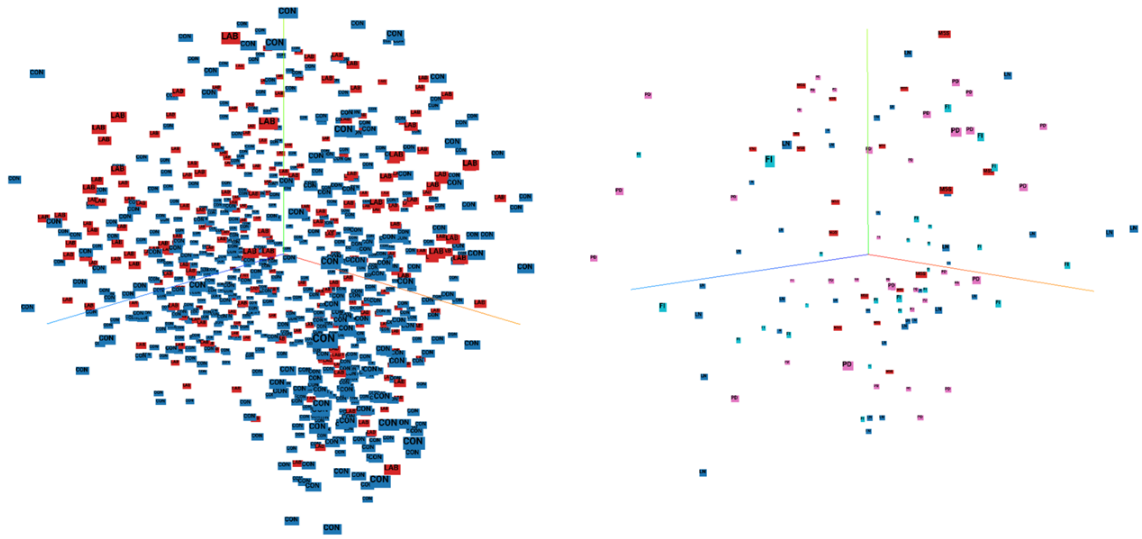

Figure 3: PCA projections of the contextual word embeddings for the words *immigration* and *immigrazione* uttered in UK's House of Commons (left) and Italy's Senate (right) in 2016. Each point in the projection is labeled with the political group of the MP who uttered word. The scarcity of Italian data is likely due to the dataset problems issues discussed in Section 2. The graphs were generated using `projector.tensorflow.org`.

is as a measure of how uniformly a certain word is used within group A. If $SP(A, A)$ is low, this indicates that the multidimensional "cloud" of embeddings is more compact, meaning members of the group tend to use the word in a similar way. Conversely, if $SP(A, A)$ is high, the cloud is more dispersed, suggesting greater variation in how the word is used among members of the group. It is important to note that the formula for SP within the same group was adjusted to exclude the distances between each embedding and itself. In datasets with few data points, these self-distances, which are always zero, can significantly skew the average downward.

Figure 4 illustrates the semantic polarity results for the word *immigration* used by the main British political parties from 2015 to 2022. Figure 5 shows the results of the same analysis, mutatis mutandis, for the Italian case.

If the hypothesis that such contextual word embeddings are able to capture political polarisation were correct, we would expect to see paler colours along the diagonal (i.e., low "polarisation" values within the parties) and darker colours at the intersections of politically distant parties (i.e., high polarisation values). Additionally, if the decade-long rise in polarisation observed by Ding et al. (2023) applied to these data as well, we would also expect to see progressively darker colours below the diagonal over the years. However, as the matrices show, no such patterns are apparent, and in some

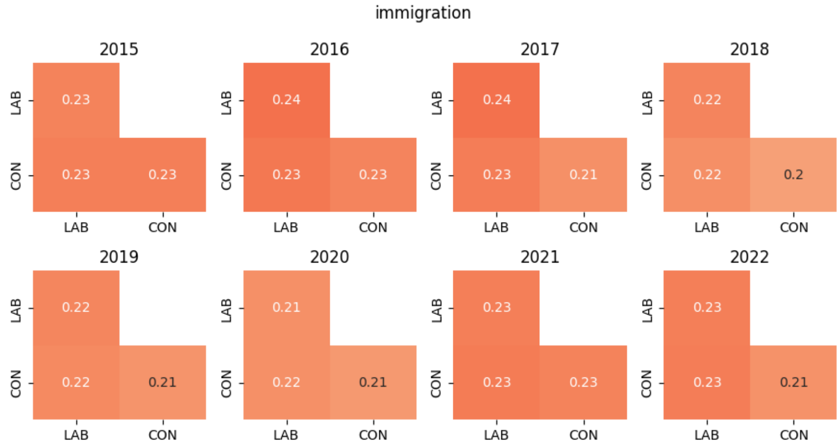

Figure 4: Semantic polarity values within and across major political parties for the use of the word *immigration* in the House of Commons from 2015 to 2022.

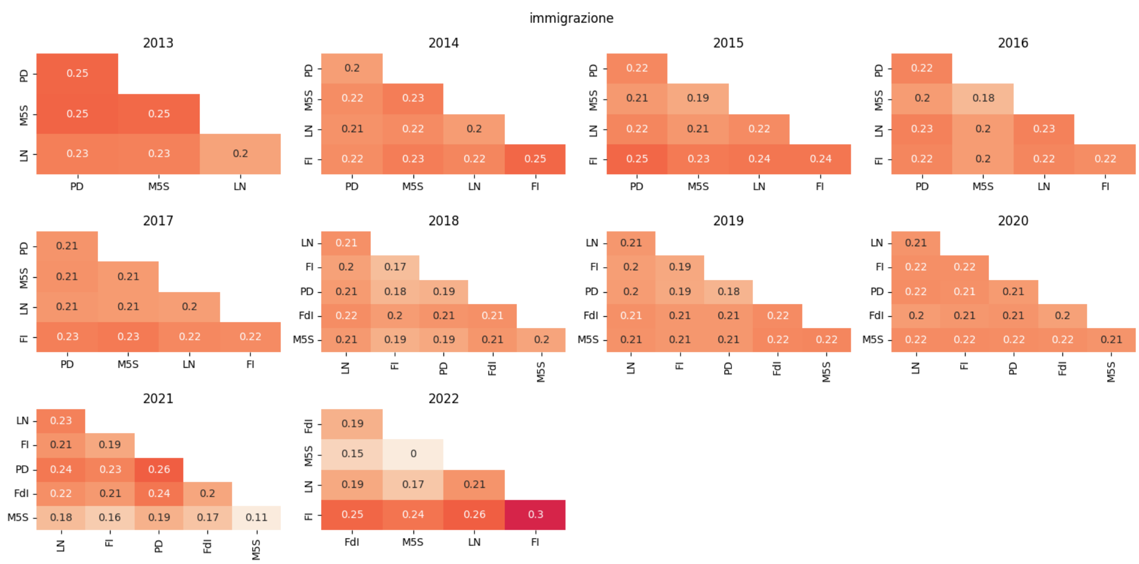

Figure 5: Semantic polarity values within and across major political parties for the use of the word *immigrazione* in the Italian Senate from 2013 to 2022.

cases, SP even peaks within the parties. This suggests that there is not clear effect of political party captured by these contextual embeddings. Results for other topical words are analogous and are available in the study's GitHub repository (see Section 7).

The slight fluctuations in SP are due to the fact that the contextual embeddings, albeit drawn from different groups, are derived from the same word form (here, *immigration*). The more pronounced fluctuations in 2021 and 2022 in the Italian results are due to fewer available data (see *immigrazione* in Figure 2), resulting in less averaging of the cosine distances.

5 Discussion

This section explores a few factors that may explain why the experiment failed on parliamentary speech, some of which were already mentioned in Sections 2 and 3. One such factor is the smaller volume of data, but the most relevant challenges are probably qualitative. Parliamentary speeches typically occur as part of a debate, where a response to a previous speaker is bound to share much of their speech's context, even when the two speakers hold very different views on the topic of discussion. This is particularly true when critique, sarcasm, mockery, or other rhetorical devices are employed, which can obscure the speaker's actual political stance. As a result, the shared vocabulary among parliamentarians may dominate, while the words that differentiate political positions may be too sparse to significantly influence the embeddings.

Another hypothesis is that parliamentary speech is highly distinctive, characterized by formal rituals and a specialized register. As such, generic mid-sized language models like `bert-base-uncased` may not be well-suited to analyse this specific form of language. Larger models may be better equipped to capture the nuances that BERT seems to fail to represent. Or yet, the formal tone of political discourse could mislead the language model's attention mechanism, causing it to prioritize grandiloquent but content-light words, while it underweights subtle cues that are of politically more meaningful.

To assess at least whether these embeddings contain any useful information for measuring polarisation, one potential approach is to train a task-specific classifier that attempts to predict the political party based on a given contextual embedding. This way, if polarisation is reflected in certain dimensions of the embeddings, the model can learn to assign higher weights to those dimensions. The classifier's performance could then serve as an indicator of polarisation, with better performance suggesting more pronounced polarization between parties. However, the low-data environment could significantly hinder such an approach, leading the model to not learn at all, to overfit, or to even "cheat", e.g. by relying on idiolectical properties of specific party members rather than trying to capture political stance.

Ultimately, it may be that contextual word embeddings are not the right tool for the purpose of measuring political polarisation. Although the experiment by Ding et al. (2023) is interesting with regard to the diachronic analysis they manage to conduct through contextual embeddings, there are several potential limitations to this method. First, as previously mentioned, contextual word embeddings encapsulate a wide range of different information, which is not just semantic, but also syntactic and possibly even pragmatic (e.g., register). Embeddings are black-box representations, making it difficult if not impossible to isolate a subset of dimensions that could be relevant for political polarisation.

Additionally, contextual word embeddings have been proven to be anisotropic (Ethayarajh, 2019), meaning they do not distribute uniformly in their vector space, which can undermine the reliability of cosine similarity measures between them. Differences in how context specificity behaves across different models (Ethayarajh, 2019) also suggest that the results of this analysis could vary significantly depending on the language models used, further complicating the analysis.

As discussed earlier in Section 1, numerous approaches to defining and measuring polarisation have been developed. Some rely on surveys, while others emply machine learning, sentiment analysis, topic modeling, or established techniques from political science, such as Wordscores, Wordfish, or Wordshoal (Németh, 2023). Some of these methods can be biased toward specific aspects of polarisation (e.g., sentiment, people's perception, etc.), while contextual word embeddings might in theory be able to capture a broad range of relevant aspects, but their black-box nature and anisotropic behaviour present significant challenges that cannot be ignored.

6 Conclusion

This study sought to apply the method proposed by Ding et al. (2023) to parliamentary speech. The lack of positive results has led to a speculation on the reasons behind the experiment's failure. Several properties of parliamentary speech data have been highlighted as possible explanations, including its debate format and formal register. Additionally, technical issues, such as data scarcity and model size, may have played a role. The study concludes that contextual word embeddings may not be suitable for this type of analysis, mainly due to their black-box nature and anisotropic behaviour.

While contextual embeddings present notable limitations, studies like the one by Ding et al. (2023) encourage further research on the topic. Regarding the purpose of measuring polarisation, however, alternative approaches – including static word embeddings – are likely more effective for now.

7 Materials

The code framework developed for this study is available on GitHub together with the semantic polarity figures for the other topical words:
`github.com/GiulioCusenza/PolarisationContextualEmbeddings`

The ParlaMint datasets used in the experiments were downloaded from the following repository hosted by CLARIN:
`clarin.si/repository/xmlui/handle/11356/1912`

References

Blumenstiel, J. E. and T. Plischke (2015). Changing motivations, time of the voting decision, and short-term volatility – the dynamics of voter heterogeneity. *Electoral Studies 37*, 28–40.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). Bert: Pre-training of deep bidirectional transformers for language understanding.

Ding, X., M. Horning, and E. H. Rho (2023). Same words, different meanings: Semantic polarization in broadcast media language forecasts polarity in online public discourse. *Proceedings of the International AAAI Conference on Web and Social Media 17*, 161–172.

Erjavec, T., M. Kopp, M. Ogrodniczuk, P. Osenova, M. Agirrezabal, T. Agnoloni, J. Aires, M. Albini, J. Alkorta, I. Antiba-Cartazo, E. Arrieta, M. Barcala, D. Bardanca, S. Barkarson, R. Bartolini, R. Battistoni, N. Bel, M. d. M. Bonet Ramos, M. Calzada Pérez, A. Cardoso, Ç. Çöltekin, M. Coole, R. Darģis, R. de Libano, G. Depoorter, S. Diwersy, R. Dodé, K. Fernandez, E. Fernández Rei, F. Frontini, M. Garcia, N. García Díaz, P. García Louzao, M. Gavriilidou, D. Gkoumas, I. Grigorov, V. Grigorova, D. Haltrup Hansen, M. Iruskieta, J. Jarlbrink, K. Jelencsik-Mátyus, B. Jongejan, N. Kahusk, M. Kirnbauer, A. Kryvenko, N. Ligeti-Nagy, N. Ljubešić, G. Luxardo, C. Magariños, M. Magnusson, C. Marchetti, M. Marx, K. Meden, A. Mendes, M. Mochtak, M. Mölder, S. Montemagni, C. Navarretta, B. Nitoń, F. M. Norén, A. Nwadukwe, M. Ojsteršek, A. Pančur, V. Papavassiliou, R. Pereira, M. Pérez Lago, S. Piperidis, H. Pirker, M. Pisani, H. v. d. Pol, P. Prokopidis, V. Quochi, P. Rayson, X. L. Regueira, A. Rii, M. Rudolf, M. Ruisi, P. Rupnik, D. Schopper, K. Simov, L. Sinikallio, J. Skubic, L. M. Tungland, J. Tuominen, R. van Heusden, Z. Varga, M. Vázquez Abuín, G. Venturi, A. Vidal Miguéns, K. Vider, A. Vivel Couso, A. I. Vladu, T. Wissik, V. Yrjänäinen, R. Zevallos, and D. Fišer (2024). Multilingual comparable corpora of parliamentary debates ParlaMint 4.1.

Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In K. Inui, J. Jiang, V. Ng, and X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 55–65. Association for Computational Linguistics.

Goffman, E. (1974). Frame analysis: An essay on the organization of experience. *Northeastern UP*.

Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly 80*(S1), 392–410.

Mair, P. (2023). *Ruling the void: The hollowing of Western democracy*. Verso books.

McCombs, M. E. and D. L. Shaw (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly 36*(2), 176–187.

Nelson, T. E., Z. M. Oxley, and R. A. Clawson (1997). Toward a psychology of framing effects. *Political Behavior 19*(3), 221–246.

Németh, R. (2023). A scoping review on the use of natural language processing in research on political polarization: trends and research prospects. *Journal of Computational Social Science*.