Proceedings of ParlaCLARIN III @ LREC2022 pages 92–100
Marseille, 20 June 2022

FrameASt: A Framework for Second-level Agenda Setting in Parliamentary Debates through the Lens of Comparative Agenda Topics

Christopher Klamm, Ines Rehbein, Simone Paolo Ponzetto
Data and Web Science Group
Mannheim University, Germany
{klamm, rehbein, ponzetto}@uni-mannheim.de

Abstract

This paper presents a framework for studying second-level political agenda setting in parliamentary debates, based on the selection of policy topics used by political actors to discuss a specific issue on the parliamentary agenda. For example, the COVID-19 pandemic as an agenda item can be contextualised as a health issue or as a civil rights issue, as a matter of macroeconomics or can be discussed in the context of social welfare. Our framework allows us to observe differences regarding *how* different parties discuss the *same* agenda item by emphasizing different topical aspects of the item. We apply and evaluate our framework on data from the German Bundestag and discuss the merits and limitations of our approach. In addition, we present a new annotated data set of parliamentary debates, following the coding schema of policy topics developed in the Comparative Agendas Project (CAP), and release models for topic classification in parliamentary debates.

Keywords: framing, agenda setting, comparative agendas project

1. Introduction

In recent years, the concept of *framing* (Bateson, 1955; Goffman, 1974; Tversky and Kahnemann, 1984) has received more and more attention in the social sciences, focussing mostly on the analysis of media communication and its impact on politics (Entman, 1993; Scheufele, 1999; Boydstun, 2013). The importance of *framing* lies in its power to shape the way in which we perceive, organize and interpret the world around us. Studies on framing have identified different types of framing effects, such as *entity framing* (Entman, 1993), i.e., the selection and highlighting of some aspects of a perceived reality, in order "to promote problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described." (Entman, 1993, p.52). Another related framing type is *agenda setting* (Iyengar and Kinder, 1987; McCombs and Reynolds, 2002, etc.), looking at how the media influences which topics succeed in gaining public attention or, for *political* agenda setting, which topics receive attention in politics (for example, by succeeding in being put on the agenda in parliament). In short, agenda setting is "more concerned with which issues are emphasized, i.e., *what* is covered, than *how* such issues are reported and discussed" (Weaver, 2007, p.142). An extension of the concept is *second-level agenda setting* or *attribute agenda setting* which –in contrast to first-level agenda setting– does not primarily consider which issues are salient in the discourse but which attributes of the issue are highlighted and how they are presented. Thus, second-level agenda setting is closely related to framing, as pointed out by Weaver (2007).

In our work, we consider both aspects, (i) *what* issues are being covered in parliamentary debates and (ii) *how* they are being discussed by different political actors and parties. We investigate this by looking at parliamentary speeches held by members of *different* political parties, but on the *same* agenda item. This setting allows us to control for topic while, at the same time, observe crucial differences in *how* parties discuss a particular topic.

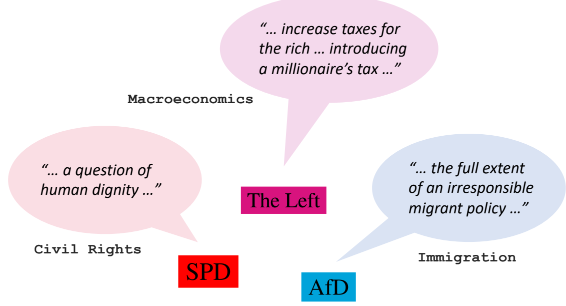

Figure 1: Example for second-level agenda setting in parliamentary debates from the German Bundestag where speakers from different parties highlight different aspects (Civil Rights, Macroeconomics, Immigration) of the same agenda item (Immigration).

For illustration, consider a parliamentary debate on the policy topic of Immigration and, more specifically, on benefits for asylum seekers. All contributions to this agenda item are expected to address the topic under discussion but might do so by emphasizing different aspects related to this issue (Figure 1).[1] One party might blame the government for their allegedly irresponsible immigration policy, another party might focus on civil rights aspects while yet another party might discuss the topic from a macroeconomic perspective by proposing a millionaire's tax. As a result, this setting provides

[1] All three excerpts are taken from a debate in the German Bundestag on 14/03/2019, Session 86, agenda item "Zusatzpunkt 6".

us with an ideal testbed that allows us to analyse and compare how parties frame the discussion of the *same* topic in different ways.

This paper proposes **FrameASt**, a **Frame**work for Second-level **A**genda **Set**ting in Parliamentary Debates. We conceptualise framing as described above and operationalise it by looking at the differences in the selection of agenda policy topics by different political actors, based on the Comparative Agendas Project (CAP) framework (Bevan, 2019) (Section 2). To identify the different CAP topics in the debates, we develop supervised CAP topic classifiers, as described in Section 3. We evaluate our classifiers on a data set of interpellations annotated within the Comparative Agendas Project and present a new data set for German parliamentary debates, annotated with major and minor CAP topics. In Section 5 we discuss applications and potential limitations of our framework before we conclude and outline avenues for future work.

2. Related Work

We first review related work on topic classification (§2.1) and framing in parliamentary debates (§2.2) and then introduce the Comparative Agendas Project (Bevan, 2019) (§2.3).

2.1. Topic Classification for Political Text

Previous work on topic classification for political text has looked at both, unsupervised and supervised approaches. Herzog et al. (2018) predict Comparative Agenda topics for debates from the UK House of Commons, using an unsupervised topic modelling approach with transfer topic labelling. Similar in spirit is the work of Brand et al. (2021) who also use the CAP codebook to detect policy agendas, however, their approach is based on Heterogeneous Information Networks (HIN) and node embeddings for identifying policy fields in German parliamentary debates. Kreutz and Daelemans (2021) is an example for a semi-supervised approach based on CAP policy agendas. Their method makes use of an automatically generated lexicon, based on graph propagation.

Many supervised approaches have been using data from the Manifesto Project database[2]. Zirn et al. (2016) predict coarse topics on the sentence level for electoral manifestos for English (Zirn et al., 2016) and (Glavaš et al., 2017) predict topics in a cross-lingual setting (Glavaš et al., 2017). Verberne et al. (2014) also work with political manifestos but predict a set of over 200 fine-grained topics on the level of semantically coherent text segments. Subramanian et al. (2018) apply deep neural networks for manifesto policy classification where they first predict the labels, based on a hierarchical multitask model, and then use probabilistic soft logic to refine them.

More recent work explores transformer-based transfer learning (Vaswani et al., 2017) for topic classification. Abercrombie et al. (2019) present a corpus of UK parliamentary debates, annotated with policy preference codes (i.e., the domain of a policy issue and the stance towards this issue) from the Manifestos Project and develop models for the automatic prediction of those topics. Koh et al. (2021) use transformers to predict labels for English political manifestos on the sentence level.

In our work, we compare a simple feature-based bag-of-words SVM classifier to transformer-based BERT models (Devlin et al., 2019), fine-tuned for the prediction of CAP topics on the level of semantically coherent segments.

2.2. Framing in Parliamentary Debates

While most work on framing has studied how the selection and presentation of topics in the mass media shapes public opinion, far fewer studies have looked at agenda setting and framing in political communication.

Naderi and Hirst (2016) study framing strategies in Canadian parliamentary debates, focussing on the recognition of a set of predefined frames on the topic of same-sex marriage. Their work can be described as entity framing in the sense of Entman (1993).

Umney and Lloyd (2018) take an original approach to framing from the perspective of design studies and investigate how political actors make use of precedents to reframe the political discourse in parliamentary debates. Otjes (2019) points out the lack of agenda setting studies in the context of parliamentary settings, due to the fact that this process usually takes place behind closed doors and thus the data needed for empirical studies is not available. An exception is the Dutch parliament, the *Tweede Kamer*, where the decision of what to put on the agenda is made in public. This allows Otjes to study how parties from the government and opposition act to promote their "own" issues[3] and, at the same time, fight other parties' attempts to do the same. The relevance of this practice has been pointed out by Otjes (2019, p.731).

> "If a party is able to set the tone in parliament, they may be able to set the themes for the election campaign."

We follow previous work (Otjes, 2019; Green-Pedersen and Mortensen, 2010) and look at agenda setting in parliament from an *issue ownership* (or *issue competition*) perspective (Walgrave et al., 2015), based on the assumption that parties have a strong preference for promoting policy issues that are associated with them and where voters assume that they are competent to deal with this issue. In contrast to Otjes (2019), we are not able to study the selection of agenda items in parliament, due to the lack of data. However, what we can investigate is *how* the different agenda items are discussed

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

[3] For a definition of *issue ownership*, see, e.g., Walgrave et al. (2015; Stubager (2018).

1	Macroeconomics	*6*	Education	*12*	Law and Crime	*17*	Technology
2	Civil Rights	*7*	Environment	*13*	Social Welfare	*18*	Foreign Trade
3	Health	*8*	Energy	*14*	Housing	*19*	International Affairs
4	Agriculture	*9*	Immigration	*15*	Domestic Commerce	*20*	Governmental Operations
5	Labor	*10*	Transportation	*16*	Defense	*21*	Public Lands
						23	Culture

Table 1: The 21 major agenda policy topics in the CAP schema (agendas 11 and 22 have been removed by the CAP project when revising and unifying the annotations from multiple participating projects).

by the different parties. As put by Otjes (2019), "Political competition focuses on selective emphasis of issues rather than direct confrontation on those issues". This *selective emphasis* is the focus of our framework, and we propose to study it by comparing the *differences* in the selection of policy topics by the different parties in debates of the *same* agenda item.

2.3. The Comparative Agendas Project

The main goal of the initial Comparative Agendas Project was to track agenda setting in the news, i.e., how much attention is being directed towards an issue. In order to achieve this, Baumgartner and Jones (1993) used distant reading techniques by looking at the headlines and abstracts of over 22,000 media articles, to identify the key topics covered in the news. This early work triggered many follow-up studies on identifying and tracking topical issues in different types of political text and for many countries (for an overview, see (Baumgartner et al., 2019)). A major contribution of project is that they make their data available to the research community, which enabled comparative studies of public policy on a large scale.

The unified schema of the CAP data (Bevan, 2019) focusses on topical issues, intentionally ignoring the framing of those issues (i.e., aspects such as positive or negative stance or ideological position). The reason behind not encoding those aspects in the CAP schema is by no means a lack of interest but rather due to the contextual sensitivity of framing that requires not only a lot of thought during coding but also in-depth knowledge of the issue at hand. This would make the large-scale annotation of text infeasible. Despite this, the CAP topics (see Table 1) provide a valuable basis for framing analysis, as we hope to show in our work.

3. CAP Topic Classification

We now present different classifiers trained to predict the 21 major CAP policy labels that we later use in our analysis. We model the identification of the policy topics as a segment-level classification task. Our first model is a feature-based SVM classifier that we compare to a transfer learning approach based on transformers (Devlin et al., 2019).

3.1. Text Segmentation

To segment the speeches into semantically coherent texts, we apply the unsupervised text segmentation algorithm of Glavaš et al. (2016) which creates a semantic relatedness graph of the input text, based on the similarity of word embeddings for words in the text. To obtain semantically coherent text segments, a graph-based segmentation algorithm then tries to find maximal cliques in the relatedness graph.

3.2. Topic Classification

We adapt the graph segmentation model to German[4] and apply it to the speeches in our corpus. We use a relatedness threshold of 0.1 and a minimal segment size of 1. The first parameter is used in the construction of the relatedness graph while the second parameter defines the minimal segment size, where 1 specifies a minimum number of one sentence per segment. The parameters have an impact on the number of segments produced by the model and have been chosen so that the segment size is reasonably large while allowing the model to split up larger, semantically unrelated text passages. During segmentation, the 25,311 speeches have been split up into 37,553 segments which are the input to our topic classifiers.

SVM Topic Classifier We train an SVM topic classifier on the Parliamentary Question Database from the CAP project[5] , a data set with more than 10,000 major and minor interpellations posed by parliamentarians (mostly from the opposition parties) to the government (Breunig and Schnatterer, 2019). The data set ranges over the 8th to the 15th legislative periods (1976–2005). Each interpellation has been assigned a major and a minor CAP topic.

We first applied some standard preprocessing and clean-up steps to the data where we also removed meta-information, such as listings of politicians' names and header/footer information. We removed stopwords and punctuation and extracted a) a tokenised and b) a lemmatised version of the data.[6] After preprocessing and clean-up, the interpellations have an average length of 388 tokens. The length range varies from 17 to 7,253 tokens per interpellation, with a standard deviation of 429 tokens.

We then trained feature-based text classification models on the preprocessed data, based on bag-of-words (BOW)

[4]The original model has been developed for English.

[5]The annotations are available from `https://www.comparativeagendas.net/datasets_codebooks`.

[6]For lemmatisation, we used the spaCy library: `https://spacy.io` with the `de_core_news_sm` model.

id	Topic	SVM	BERT	ParlBERT	support
19	International	77.4	78.7	**80.0**	1,126
16	Defense	84.0	**85.0**	**85.0**	1,099
20	Government	66.1	69.6	**71.3**	989
2	Civil Rights	**79.3**	78.8	76.5	978
7	Environment	76.3	76.3	**76.6**	845
10	Transportation	83.8	**87.7**	86.0	800
12	Law & Crime	66.3	65.7	**67.1**	492
8	Energy	76,2	76.0	**78.6**	424
3	Health	76.8	**82.3**	78.2	418
15	Domestic Com.	57.1	**66.6**	64.4	382
9	Immigration	74.8	80.3	**81.0**	376
5	Labor	67.9	**70.0**	69.1	344
1	Macroeconom.	61.5	61.1	**62.8**	339
4	Agriculture	77.9	**78.7**	76.3	292
13	Social Welfare	**55.1**	54,1	49,2	253
17	Technology	58.7	**67.8**	63.0	252
6	Education	64.0	67.6	**71.6**	183
14	Housing	73.0	78.5	**79.6**	178
18	Foreign Trade	51.0	58.1	**61,5**	139
23	Culture	**68.8**	64.2	54.6	69
21	Public Lands	32.4	42.4	**45.4**	55
	total f1-micro	73.7	76,8	76.5	10,033

Table 2: Results (micro F1) for different classifiers (SVM, GermanBERT (BERT) and GermanParlaBERTarian (ParlBERT)) for CAP topics. Support shows the number of training instances for each class.

features with and without tf-idf weighing. We experimented with different classification algorithms from the scikit-learn library[7], where the linear SVM achieved best results. Hyperparameters have been determined in a 5-fold cross-validation setup (20k features, w/o tf-idf on lemmatised unigrams). Other parameters have been set to *penalty: l2, loss: squared hinge, C: 1.0, max_iter: 1,000*.[8]

Our best SVM classifier achieves a micro F1 over all 21 agenda topics of 73.7% on the in-domain interpellation data (Table 2). Results for the individual classes range from 32.4 to 84.0, with higher scores for the more frequent labels, as is common for supervised machine learning models. The 10 policy topics with the highest F1 scores are Defense (84%), Transportation (84%), Civil Rights (79%), Agriculture (78%), Health (77%), International Affairs (77%), Energy (76%), Environment (76%), Immigration (75%) and Housing (73%).

GermanParlaBERTarian (ParlBERT) To obtain topic predictions, we first adopted an existing GermanBERT [9] language model with adaptive pre-training on German parliamentary debates from the DeuParl corpus (Walter et al., 2021). This domain adaptation step is a form of transfer learning, with the goal of adapting a model trained on a *source domain*, for example Wikipedia, to a *target domain*, such as parlamentary debates (Ruder, 2019).

The data we used to perform this adaptation includes sentences from different decades and legislative terms, spanning a time period from 1867-2020, with a broad range of speakers from all political parties in Germany. We used unsupervised masked language modelling for two epochs with a learning rate of $5e-5$ and a batch size of 16 on more than 8 million sentences with a minimum sequence length of 250. Then we fine-tuned the model on the CAP interpellation data for topic prediction. We trained the model for 5 epochs, with a learning rate of $5e-5$ and a batch size of 16. All experiments were averaged across ten runs with different splits. The splits as well as the models will be made publicly available.

Table 2 shows the performance of GermanBERT without domain adaption (f1-micro 76.8%) and ParlBERT with domain adaption (f1-micro 76.5%). Unlike the positive effects of domain adaptation reported for other NLP tasks (Beltagy et al., 2019; Lee et al., 2020), we did not see any substantial improvements but observed results in the same range as for GermanBERT with task-specific fine-tuning. However, compared to GermanBERT, ParlBERT seems to yield higher results on low-frequency topics.

While the BERT-based models generally outperform the SVM classifier, for some topics the SVM achieves higher results (e.g., Civil Rights, Energy, Social Welfare, Culture). This indicates that the performance is highly topic-dependent. Overall, GermanBERT and ParlBERT show a promising performance for CAP topic classification while, at the same time, the differences in results across topics illustrate the computational challenges for topics with small sample sizes and the overall need for more research in the area of few-shot learning.

4. Data and Annotation

The data we use in our work are parliamentary debates from the German Bundestag (mostly) from the 19th legislative period (Oct 24, 2017 to Nov 18, 2021). Our corpus includes over 14 mio. tokens from speeches held by 759 different speakers (Table 3).[10] [11]

4.1. Sampling

From this corpus, we selected a sample of speeches for manual annotation. Our objective was to create a gold standard controlled for topic, with roughly the same amount of text for each party. To obtain our goal, we sampled the data as follows.

First, we identified agenda items from the Bundestag debates that covered different policy topics in the CAP

[7] https://scikit-learn.org

[8] We also experimented with a larger number of (unigram and ngram) features and with other algorithms from the scikit-learn library but obtained best results for the linear SVM with unigram features.

[9] https://huggingface.co/bert-base-german-cased with HuggingFace (Wolf et al., 2020)

[10] The abbreviations in Table 3 refer to the *Christian Democratic Union* in Germany and *Christian Social Union* in Bavaria CDU/CSU; the *Social Democratic Party* SPD; the *Alternative for Germany* AfD; the *Free Democratic Party* FDP; the *Greens* The Greens and *The Left* The Left.

[11] We removed 84 speeches for which the given speaker information was not sufficient to unambiguously identify the party affiliation (e.g., *Roth* could refer to Michael Roth (SPD) or to Claudia Roth (Greens).

party	# speeches	# tokens	# spk
CDU/CSU	7,635	4,862,654	259
SPD	5,321	3,158,315	167
AfD	3,465	1,844,707	95
FDP	3,067	1,593,108	89
The Greens	2,866	1,522,305	70
The Left	2,671	1,394,089	72
cross-bencher	200	86,170	7
total	25,225	14,461,348	759

Table 3: Some statistics for our corpus of Bundestag debates (token counts excluding punctuation). *Cross-bencher* refers to members of the parliament not affiliated with any political party.

schema. The identification of agenda items was based on the supervised SVM classifier described in Section 3. We used the model to predict major CAP policies for all speeches in our data and then assigned the majority label to the agenda item, to determine the *main topic* for this item. For illustation, let us assume that we have an agenda item i on some topic t and we have all the parliamentary debate contributions by politicians from different parties on this particular agenda item. Let us also assume that we have 10 debate contributions for this agenda item and that our classifier predicted the major CAP policies "Immigration" for 6 of the 10 speeches and "Civil Rights" and "Social Welfare" for the remaining four speeches (see Table 4 below). Then the majority label for this agenda item, i.e., its *main topic*, would be "Immigration".

Agenda item i with 10 speeches										
Speech:	1	2	3	4	5	6	7	8	9	10
Topic:	I	I	C	I	S	C	I	I	C	I
Majority label for i: I (Immigration)										

Table 4: Example for determining the majority topic for an agenda item i with 10 speeches. *Topic* refers to the CAP topic predicted by the SVM. I, C and S stand for **I**mmigration, **C**ivil Rights and **S**ocial Welfare.

Based on the majority labels for agenda items, we identified relevant agendas for each major CAP policy label from which we then selected and manually validated one agenda item for manual annotation, based on the following criteria: (a) we select agenda items that include speeches by members from each of the 6 parties and (b) we select agenda items where at least 60% of the predictions made by the classifier agree on a topic (which is what we call the *main topic* of the agenda item). However, we do not select items where all or nearly all (i.e., 80% or more) of the predictions agree on the topic, as we want to avoid creating an unrealistically "easy" validation set.
Following this procedure, we extracted a validation set

party	# speeches	# tokens	# spk
CDU/CSU	57	37,636	47
SPD	45	26,124	37
AfD	25	14,514	22
FDP	22	12,466	19
The Greens	25	13,574	18
The Left	22	12,295	19
cross-bencher	1	284	1
total	197	116,893	163

Table 5: Some statistics for our manually annotated test set (token counts excluding punctuation).

for manual annotation with more than 100,000 tokens and with 7 to 14 speeches per agenda item (see Table 5). The advantage of our sampling procedure is that it allows us to compare speeches by political actors from different parties on exactly the same topic (i.e., agenda item) and to investigate which aspects of this agenda item have been emphasized by each party.

4.2. Annotation

The CAP coding schema includes 21 major topics and more than 200 fine-grained subtopics. We follow the CAP schema and annotate major and minor topics in our data set, to be used for an in-domain evaluation of our classifiers.

Annotation Process The annotators are two NLP researchers with experience in linguistic annotation but have not worked with the CAP schema before.[12] They were presented with the speeches, one at a time, and were instructed to first read through the whole text of the speech. Then they segmented each speech into semantically coherent text segments, based on the policy topics discussed in the text, and assigned one major and minor CAP topic label to each text segment. The annotators were instructed to introduce new segment boundaries only if they noticed a change in topic.
For annotation, we split our data into three batches. The first batch included one speech only for each major CAP label, to familiarize the annotators with the relevant topics for this agenda item. The second batch was considered as a training round where each speech was annotated independently by each of the annotators. The third batch has been annotated after the completion of the training round and reflects the quality of the annotatons. After each round of annotation, all disagreements have been resolved in discussion.

Inter-Annotator Agreement We now report results for inter-annotator agreement (IAA) for the 21 major CAP topics for the second (training round) and third batch of labelled debates in our new dataset. We collect the set of all CAP labels that have been assigned to a specific speech and compare the sets of labels assigned by the two annotators. As our data includes multiple

[12]The first two authors of the paper.

labels per speech, we cannot compute Cohen's kappa or related measures. Instead, we report the Jaccard similarity between the two sets of assigned labels for each speech.
Given two sets of labels, A and B, we compute the Jaccard similarity coefficient for each speech in our data as shown below (Equation 1).[13]

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

The average over the Jaccard similarity coefficients for all speeches in the second batch (training round) is 0.59, for the third batch the score increases to 0.74.

Challenges for Annotation Both annotators found the task challenging, due to the scarse guidelines in the codebook which only presented the annotators with minimal descriptions of each CAP policy topic but did not include a more theoretical discussion on how to distinguish between related and overlapping topics. Other challenges for annotation were posed by the identification of the exact segment boundaries. Here the two annotators often identified a change in topic but did not agree on the exact point of segmentation (i.e., topic change).

4.3. Topic Classification on Parliamentary Debates from the German Bundestag

We now evaluate our topic classifiers on the newly created data set from the German Bundestag. After the manual annotation step had been completed, we mapped the annotated major and minor CAP topic labels onto the automatically created text segments (see Section 3.1), to create a new data set of parliamentary debates with major and minor CAP topic labels on the segment level. This procedure can result in more than one gold label per segment in cases where the human annotators decided on a topic change for segments that have not been split by the text segmentation algorithm. As our classifier can only predict one label per segment, we decided on a lenient evaluation strategy that does not punish the classifier for non-optimal segmentation decisions. Our procedure is as follows: For each text segment, we count the predicted label as a true positive (TP) if it is included in the set of manually assigned labels for this segment, and as a false positive (FP) otherwise. We then report the accuracy for each topic and micro F1 over all topic classes.
Table 6 reports results for the three classifiers on the segments from our new dataset of parliamentary debates from the Bundestag, annotated for CAP topics. As expected, results are a bit lower than for the in-domain interpellation data. This might reflect an out-of-domain effect for the debates. It is also conceivable that our interpretation of the CAP guidelines was slightly different than the one of the original annotators. Another possible explanation for the lower results is the automatic segmentation process which might not always yield optimal results.
Overall, we observe a similar trend as for the interpellations data (Table 2), with lower results for the less frequent classes (such as Culture, Foreign Trade, Technology, Social Welfare). A bit surprising is the decrease in results for some of the more frequent classes (Government Operations, Environment, Transportation).
Taking a look at the data, we notice that for the "Environment" class, only 4 out of 10 instances in the ParlaBERT results have been predicted correctly. Four of the incorrect predictions have been annotated as "Energy" in the gold data set while the remaining two cases were labelled as "Agriculture". For "Transportation", on the other hand, 12 of the 14 incorrect predictions have been annotated as "Energy" by the human coders and another one as "Environment". This reflects the close thematic interconnection of the three topics, Energy, Transport and Environment, in recent parliamentary debates which poses a challenge for CAP topic classification.

id	**Topic**	**SVM**	**BERT**	**ParlBERT**
19	International	28.1	86.7	75.0
16	Defense	42.9	**100.0**	85.7
20	Government Operations	34.3	33.3	**44.8**
2	Civil Rights	63.0	90.0	**94.7**
7	Environment	38.5	28.6	**40.0**
10	Transportation	43.8	**58.3**	33.3
12	Law and Crime	61.5	83.3	**90.0**
8	Energy	90.9	**100.0**	**100.0**
3	Health	**100.0**	**100.0**	90.0
15	Domestic Commerce	**88.9**	58.8	70.6
9	Immigration	**100.0**	**100.0**	92.9
5	Labor	72.2	95.0	**95.2**
1	Macroeconomics	**100.0**	90.0	**100.0**
4	Agriculture	**100.0**	**100.0**	94.4
13	Social Welfare	60.0	**71.4**	40.0
17	Technology	60.0	**66.7**	50.0
6	Education	**57.1**	25.0	0
14	Housing	**100.0**	**100.0**	**100.0**
18	Foreign Trade	0	0	0
23	Culture	0	0	0
21	Public Lands	n/a	n/a	n/a
	total f1-micro	58.3	68.7	**70.2**
	(true positives)	(147/252)	(173/252)	(177/252)

Table 6: Results (micro F1) for the CAP topic classification models (SVM, (German)BERT and ParlBERT) on the newly annotated data set of parliamentary debates (252 segments).

5. Potential Applications and Limitations

We now summarise the main components of our framework and discuss potential applications as well as limitations of our work.
Figure 2 illustrates the different components of our pipeline. Given **(1)** a set of parliamentary speeches, we **(2)** predict CAP topics for all speeches in our parliamentary data, using our supervised topic classifier. Next, we group all speeches that belong to the *same*

[13]The Jaccard similarity for two identical sets is 1.0. A comparison of two sets where the second set is a subset half the size of the first set (e.g., $s_1 = [1, 2, 3, 4]$ and $s_2 = [1, 4]$) would yield a Jaccard similarity of 0.5.

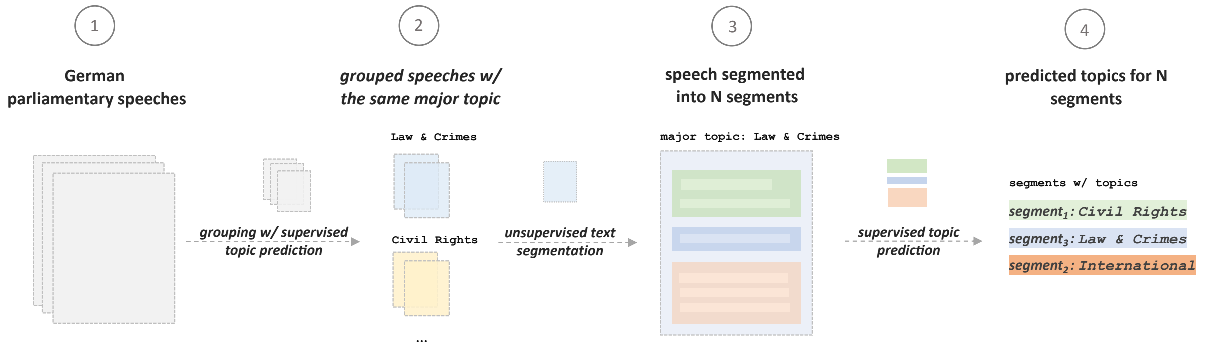

Figure 2: Overview of our framework illustrating the workflow for a given **(1) set of parliamentary speeches (2) grouped along shared topics** and **(3) applied segmentation** with **(4) supervised topic prediction** for all identified segments.

agenda item and determine the dominant topic for this set of speeches, based on the majority label predicted by our CAP topic classifier. We only include topics with a prediction accuracy of at least 75% (Defense (85%), Transportation (88%), Civil Rights (79%), Agriculture (79%), Health (82%), International Affairs (79%), Energy (76%), Environment (76%), Immigration (80%) and Housing (78%)). We then use **(3)** the unsupervised text segmentation model of Glavaš et al. (2016) and split the speeches into semantically coherent text segments. In the next step, **(4)** we use our topic classifier to predict CAP topics for each *speech segment*, which results in a set of semantically coherent, topic-annotated speech segments for each agenda item.

Our framework provides us with the means for comparing *how* different parties discuss the same *main topic* (or dominant topic, based on the majority predictions of the CAP topic classifier) of an agenda item, i.e., *which topics* are emphasized by each party in the plenary debates. In addition, it allows us to track the salience of specific topics over time. We plan to use our methodology to study the emergence of new topics over a longer period of time (e.g., climate change) and whether and how they have been adopted by political actors in the parliamentary setting.

Figure 3 shows a prototypical use case of our framework. It illustrates the distribution of CAP topics (e.g., Defense, Transportation, International Affairs, etc.) that have been used by the different German parties (SPD, CDU/CSU, AfD, The Left, The Greens and the FDP) when discussing *the same* main topic. For example, for the debates of all agenda items that have been predicted a certain majority topic, *which other topics have been used by members of the different parties in debates of this particular main topic*?

Overall, we can observe that the normalized distribution of topics across all parties follows a slightly different pattern. We can identify topics that seem to be more associated with certain parties. For instance, the CAP topic "Civil Rights" is more often emphasized in debate contributions by the Left, the Greens, and

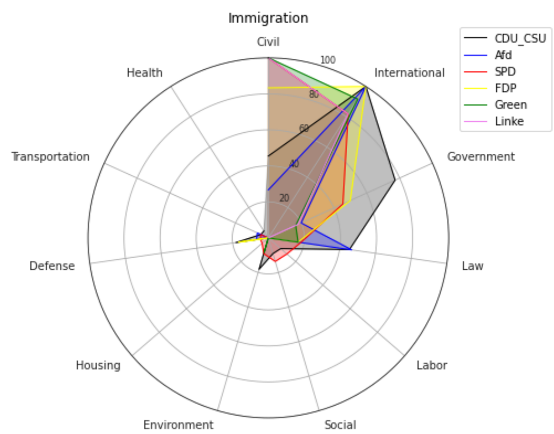

Figure 3: Normalized distribution of the associated topics per party AfD, SPD, CDU/CSU, The Left, FDP and The Greens in Germany regarding the major topic Immigration.

the SPD. In contrast, the AfD shows a below-average use of the "Civil Rights" topic in their speeches when talking about immigration, putting a stronger focus on "Law and Crime". In comparison, the CDU/CSU seems to associate the topic more often with aspects related to "Government Operations". This examaple should give the reader a first idea of possible applications and research questions that could be studied with our framework for second-level agenda setting in parliamentary debates.

There are also limitations to our work. In particular, our framework only allows us to investigate *which policy issues* have been emphasized in the debates, but not the stance of a particular party towards this issue. For example, two parties might emphasize the same policy issue but might still persue diametrically oposed interests. One straightforward way to address this issue is the extension of our framework with topic-based (or issue-based) stance detection. We plan to persue this avenue in future work.

6. Conclusion

In the paper, we introduced a framework for the analysis of plenary debates, with a focus on second-level agenda setting. Our framework allows us to observe differences in how political parties discuss the same policy issues by highlighting different thematic aspects of the issue. We have applied our framework to data from the German Bundestag and contribute a new annotated dataset of parliamentary debates. Our annotation experiment shows the challenges for topic annotation in political debates and the computational challenges for topic classification for datasets with unbalanced and small sample sizes. We hope that our new corpus will serve as a way to better understand the variety of topic aspects associated with an agenda item in political debates.[14]

7. Acknowledgements

This work was supported in part by the SFB 884 on the Political Economy of Reforms at the University of Mannheim (projects B6 and C4), funded by the German Research Foundation (DFG), and by the Ministry of Science, Research and the Arts Baden Württemberg (MWK).

8. Bibliographical References

Abercrombie, G., Nanni, F., Batista-Navarro, R., and Ponzetto, S. P. (2019). Policy preference detection in parliamentary debate motions. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 249–259, Hong Kong, China, November. Association for Computational Linguistics.

Bateson, G. (1955). A theory of play and fantasy. *Psychiatric Research Reports*, 2:39–51.

Baumgartner, F. R. and Jones, B. D. (1993). *Agendas and instability in American politics*. Chicago: University of Chicago Press.

Baumgartner, F. R., Breunig, C., and Grossman, E. (2019). *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.

Beltagy, I., Lo, K., and Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China, November. Association for Computational Linguistics.

Bevan, S. (2019). Gone fishing: The creation of the Comparative Agendas Project master codebook. In Frank R. Baumgartner, et al., editors, *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.

Boydstun, A. E. (2013). *Making the news: Politics, the media, and agenda setting*. Chicago: U. of Chicago Press.

[14]The dataset and topic classifiers are available for download: https://github.com/chkla/FrameASt.

Brand, A., Schünemann, W. J., König, T., and Preböck, T. (2021). Detecting policy fields in German parliamentary materials with heterogeneous information networks and node embeddings. In *The 1st Workshop of Computational Linguistics for Political Text Analysis (CPSS)*.

Breunig, C. and Schnatterer, T. (2019). Policy agendas in Germany. In Frank R. Baumgartner, et al., editors, *Comparative Agendas Project: Theory, Tools, Data*. Oxford: Oxford University Press.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, et al., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43:51–58.

Glavaš, G., Nanni, F., and Ponzetto, S. P. (2016). Unsupervised text segmentation using semantic relatedness graphs. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 125–130, Berlin, Germany, August. Association for Computational Linguistics.

Glavaš, G., Nanni, F., and Ponzetto, S. P. (2017). Cross-lingual classification of topics in political texts. In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 42–46, Vancouver, Canada, August. Association for Computational Linguistics.

Goffman, E. (1974). *Frame analysis*. New York: Free Press.

Green-Pedersen, C. and Mortensen, P. B. (2010). Who sets the agenda and who responds to in the Danish parliament? A new model of issue competition and agenda-setting. *European Journal of Political Research*, 49(2):257–281.

Herzog, A., John, P., and Mikhaylov, S. J. (2018). Transfer topic labeling with domain-specific knowledge base: An analysis of UK house of commons speeches 1935-2014. *CoRR*, abs/1806.00793.

Iyengar, S. and Kinder, D. R. (1987). *News that matters: Television and American opinion*. Chicago: University of Chicago Press.

Koh, A., Boey, D. K. S., and Béchara, H. (2021). Predicting policy domains from party manifestos with bert and convolutional neural networks. In *The 1st Workshop of Computational Linguistics for Political Text Analysis (CPSS)*.

Kreutz, T. and Daelemans, W. (2021). A semi-supervised approach to classifying political agenda issues. In *The 1st Workshop of Computational Linguistics for Political Text Analysis (CPSS)*.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

McCombs, M. and Reynolds, A. (2002). News influence on our pictures of the world. In J. Bryant et al., editors, *Media Effects: Advances in Theory and Research*, pages 1–18. Mahwah: LEA.

Naderi, N. and Hirst, G. (2016). Argumentation mining in parliamentary discourse. In Matteo Baldoni, et al., editors, *Principles and Practice of Multi-Agent Systems*, pages 16–25, Cham. Springer International Publishing.

Otjes, S. (2019). No politics in the agenda-setting meeting: plenary agenda setting in the Netherlands. *West European Politics*, 42(4):728–754.

Ruder, S. (2019). *Neural Transfer Learning for Natural Language Processing*. Ph.D. thesis, National University of Ireland, Galway.

Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1):103–122.

Stubager, R. (2018). What is issue ownership and how should we measure it? *Political Behaviour*, 40:345–370.

Subramanian, S., Cohn, T., and Baldwin, T. (2018). Hierarchical structured model for fine-to-coarse manifesto text analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1964–1974, New Orleans, Louisiana, June. Association for Computational Linguistics.

Tversky, A. and Kahnemann, D. (1984). Choices, values, and frames. *American Psychologist*, 39(4):341–350.

Umney, D. and Lloyd, P. (2018). Designing frames: The use of precedents in parliamentary debate. *Design Studies*, 54:201–218.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Verberne, S., Dhondt, E., van den Bosch, A., and Marx, M. (2014). Automatic thematic classification of election manifestos. *Information Processing & Management*, 50(4):554–567.

Walgrave, S., Tresch, A., and Lefevere, J. (2015). The conceptualisation and measurement of issue ownership. *West European Politics*, 38(4):778–796.

Walter, T., Kirschner, C., Eger, S., Glavas, G., Lauscher, A., and Ponzetto, S. P. (2021). Diachronic analysis of German parliamentary proceedings: Ideological shifts through the lens of political biases. In J. Stephen Downie, et al., editors, *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2021, Champaign, IL, USA, September 27-30, 2021*, pages 51–60. IEEE.

Weaver, D. H. (2007). Thoughts on agenda setting, framing, and priming. *Journal of Communication*, 57(1):142–147.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October. Association for Computational Linguistics.

Zirn, C., Glavaš, G., Nanni, F., Eichorts, J., , and Stuckenschmidt, H. (2016). Classifying topics and detecting topic shifts in political manifestos. In *The International Conference on the Advances in Computational Analysis of Political Text*.