A New Resource and Baselines for Opinion Role Labelling in German Parliamentary Debates

Ines Rehbein, Simone Paolo Ponzetto
University of Mannheim
ines.rehbein@uni-mannheim.de, simone.ponzetto@uni-mannheim.de

Abstract

Detecting opinions, their holders and targets in parliamentary debates provides an interesting layer of analysis, for example, to identify frequent targets of opinions for specific topics, actors or parties. In the paper, we present GEPADE-ORL, a new dataset for German parliamentary debates where subjective expressions, their opinion holders and targets have been annotated. We describe the annotation process and report baselines for predicting those annotations in our new dataset.

Keywords: Opinion Role Labelling, Political Text Analysis, Holder and Target Extraction

1. Introduction

Recent work in the area of political text analysis has seen an increasing interest in using NLP methods to investigate the sentiment and positions of political actors in parliamentary debates (see Abercrombie and Batista-Navarro (2020) for an overview). Most work, however, sticks to rather coarse-grained analyses like the prediction of sentiment (positive, neutral, negative) at the level of sentences or documents (Proksch et al., 2019; Abercrombie and Batista-Navarro, 2018) or the prediction or scaling of ideology on a binary scale (*left–right*) (Laver et al., 2003; Slapin and Proksch, 2008).

We thus argue that more work is needed to enable analyses of political text on a more fine-grained level. One possible approach is Opinion Role Labelling (ORL), i.e., the extraction of opinion holders and their targets from text. ORL offers an interesting layer of analysis by distinguishing different perspectives expressed in a text. For illustration, see Fig. 1 and the examples below.

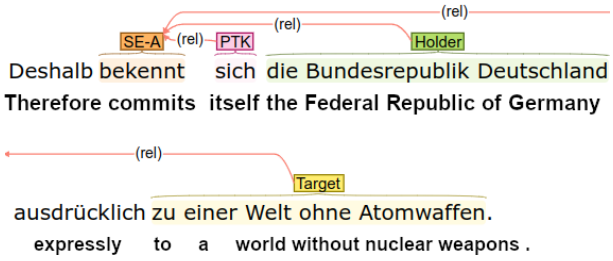

Figure 1: Example annotation from our corpus (SE-A: Subjective Expression, Agent-view; PTK: particles and reflexive pronouns).

Ex. 1.1 *The German government regrets sending the wrong message to authoritarian leaders.*

Ex. 1.2 *The German government risks sending the wrong message to authoritarian leaders.*

While both sentences express negative sentiment, the first one is written from the point of view of the German government, while the second sentence reflects the speaker's perspective. This subtle but crucial difference results in very different analyses. Instead of classiying both sentences as *negative*, a more informative analysis should capture that the first example expresses the regrets of an opinion holder (*the German government*) about an action (*sending the wrong message to authoritarian leaders*), where we can infer that the stance of the holder towards the target is negative. For the second example, we would like to know that the German government is *not* the opinion holder but the target of the opinion, while the holder is not stated explicitly but can be inferred as the speaker of the utterance.

In the paper, we present a new dataset of parliamentary debates from the German Bundestag where such differences are encoded on the level of subjective expressions (SEs) and their opinion roles. Our annotation follows a lexico-semantic approach to the identification of opinions and their holders and targets (Wiegand and Ruppenhofer, 2015), based on the detection of subjective expressions for *agent, patient* and *speaker view* verbs (for details, see Section 3). We then use our new dataset to train a state of the art Semantic Role Labelling (SRL) system that can automatically predict subjective expressions and opinion roles in text and present baselines for our new corpus.

The paper is structured as follows. We start with a short review of related work on sentiment and stance detection in political communication (§ 2) and present our lexico-semantic approach to opinion role labelling (§ 3). Section 4 describes

ParlaCLARIN IV Workshop on Creating, Analysing, and Increasing Accessibility of Parliamentary Corpora, pages 163–170

our new dataset and annotation process, and we report baselines for the automatic prediction of opinion roles in Section 5. Section 6 concludes and outlines future work.

2. Related Work

Detecting politicians' positions towards certain policy issues is an active field of research in the computational political science community (Subramanian et al., 2017; Rauh, 2018; Abercrombie and Batista-Navarro, 2018; Abercrombie et al., 2019; Koh et al., 2021; Abercrombie and Batista-Navarro, 2022).[1] However, due to a lack of resources for fine-grained analyses of the sources and targets of opinions in political debates, many works have tried to approximate the stances of political actors with sentiment predictions, assuming that the concepts are sufficiently correlated (Jose and Chooralil, 2015; Murthy, 2015; Rezapour et al., 2017; Uthirapathy and Sandanam, 2023).

Bestvater and Monroe (2023) address this issue and present three case studies showing that approximating stance with sentiment introduces noise and can thus have a negative impact on the validity of the results. Therefore, they discourage the use of sentiment dictionaries and classifiers for modelling stance and, instead, recommend to train in-domain stance classifiers for the task at hand. Below, we explain the difference between stance detection and opinion role labelling and shortly overview relevant work in each field.

Stance detection for political text analysis In contrast to sentiment classifiers that label a text as either *positive, negative* or *neutral* without specifying the target of the sentiment, a stance detection classifier takes a text and a given target and tries to determine the stance of the text toward that target as either *in favour, against* or *neither*.[2]

Work on the intersection of NLP and political science often tries to predict political preferences for a large set of fine-grained issues (Subramanian et al., 2017; Abercrombie and Batista-Navarro, 2018; Abercrombie et al., 2019; Koh et al., 2021; Abercrombie and Batista-Navarro, 2022), *inter alia*. Most notably is the Manifesto Project[3] which has created a large, multilingual collection of political manifestos across countries, where policy issues and preferences are coded on the sentence level.

Vamvas and Sennrich (2020) present a multilingual, multi-target dataset for online political debates. Mascarell et al. (2021) release a corpus of German news articles with stance annotations for a set of 91 target issues. Barriere et al. (2022a,b) create a multilingual, multi-target dataset of online debates with self-rated comments in 26 European languages. They augment the data with around 1,200 comments in 6 languages, manually annotated for stance. Göhring et al. (2021) present the German deInStance corpus, including 1,000 answers by politicians taken from the X-Stance corpus of Vamvas and Sennrich (2020), focussing on the challenging task of inferring *implicit* stances from text.

Opinion Role Labelling is the task of identifying subjective expressions in text, together with their holders and targets. Previous work has used the term "fine-grained entity or aspect-level sentiment analysis" for identifying the sentiment (*positive, negative*) of a text toward the target of an opinion (Liu, 2012), which is very similar to our goal. However, unlike aspect-level sentiment analysis and stance detection, ORL does not require any prior knowledge of the target(s), but attempts to identify them "on the fly", together with their sources.

Following the seminal work of Stoyanov et al. (2004) and Wiebe et al. (2005a) for English, Ruppenhofer et al. (2014, 2016) have presented a corpus of Swiss-German parliamentary debates annotated with subjective expressions, their opinion holders and targets. The data set has been used in two shared tasks.[4] While being similar in spirit to our work, their data is substantially smaller with around 26,500 tokens compared to over 200,000 tokens in our data. However, due to the full text annotation approach where all subjective verbs, nouns, adjectives and multi-word expressions have been coded, the density of annotated SEs in the shared task data is much higher than in our corpus.

Other work from the area of Argumentation Mining has focussed on German newswire, presenting a dataset of German newspaper articles, manually annotated for claims about the migration crisis (Lapesa et al., 2020). The authors identify and code claims, together with their holders (the ones who stated the claim), and also annotate the polarity of the claim. This results in a high-quality dataset for this particular topic. However, the approach offers limited generalisability, as the data is tailored toward one particular policy issue.

Instead, the ORL approach is more generalisable as it can be used on any text, without a predefined topic or target. This, however, comes at the cost of interpretability. While stance detection asks what stance a text conveys towards the target (e.g.,

[1] Also see Abercrombie and Batista-Navarro (2020) for a survey of recent work on sentiment and stance detection in parliamentary debates.

[2] Often the label *neutral* is also included.

[3] https://manifestoproject.wzb.eu

[4] See the IGGSA 2014 shared task: https://sites.google.com/site/iggsasharedtask/task-1 and for 2016: https://iggsasharedtask2016.github.io.

A	(Wir)$_{Holder}$	lehnen (diesen Antrag)$_{Target}$ ab$_{Ptc}$
	(We) Agent	reject (this motion) Patient
P	(Die USA)$_{Target}$	haben (mich)$_{Holder}$ enttäuscht
	(The USA) Agent	disappointed (me) Patient
S	(Deutschland)$_{Target}$	verfehlt (seine Ziele)$_{Other}$
	(Germany) Agent	fails to meet (its targets)

Table 1: Examples for agent (A), patient (P) and speaker (S) view verbs and the mapping to opinion holder and target (A: agent=holder, patient=target; P: agent=target, patient=holder; S: agent=target, holder=speaker).

a political actor like *Trump* or *Obama* or a topic like *abortion, death penalty*), the targets identified in ORL can be very heterogeneous, making it hard to map them to a predefined topic (e.g., *sending the wrong message to authoritarian leaders*). In addition, ORL does not encode the polarity of the subjective expression. The different approaches are therefore not equally suitable for all types of analyses, but should be carefully selected depending on the research question.

3. Agent, Patient and Speaker Views

To create a corpus annotated for subjective expressions, their holders and targets, we follow the lexico-semantic approach described in Wiegand and Ruppenhofer (2015). The authors show that semantic roles like *agent* and *patient* are not sufficient for distinguishing opinion *holders* from their *targets* and propose to categorise opinion verbs into three distinct views: (i) agent view, (ii) patient view, and (iii) speaker view verbs.[5]

The three views specify how the opinion holder is mapped to high-level semantic roles on the syntax-semantics interface: In the agent view, the opinion holder is the syntactic subject of the clause and is linked to the semantic role of the agent. For patient view, the holder of the opinion is not the subject but the direct object of the clause and can be mapped to the semantic role of the patient while the agent role encodes the opinion target (see Table 1). For speaker view, the semantic agent role again encodes the opinion target while the opinion holder is implicit and can be inferred as the speaker of the utterance.

Therefore, determining the correct view of the subjective expression should help us to identify the correct target as either the grammatical subject or the object of the utterance. We use this schema to create a dataset of German parliamentary debates where we annotate subjective expressions, their holders and targets and some additional roles (see Section 4). In the next section, we present our new dataset and describe the annotation process.

[5]Speaker view verbs have previously been described by Wiebe et al. (2005b) as *expressive subjectivity* and by Maks and Vossen (2011) as *speaker subjectivity*, see Wiegand and Ruppenhofer (2015).

4. Data and Annotation

Our dataset, GEPADE-ORL, includes German parliamentary debates, manually annotated for verbal subjective expressions and their opinion roles, i.e., their opinion *holders* and *targets*. The speeches are taken from the 19th legislative term of the German Bundestag, however, the distribution of topics in GEPADE-ORL is not representative of the larger data but has been sampled to cover a more diverse range of topics, with contributions from all parties distributed over the whole legislative term. Below, we describe the sampling procedure in more detail.

Sampling procedure We extracted a sample of parliamentary debates from the German Bundestag, covering all speeches from the 19th legislative term (2017–2021). The sample includes speeches by 807 different speakers, with over 900,000 sentences and over 16 mio tokens. From this corpus, we selected individual speeches for annotation, controlled for topic and including speeches for each of the political parties. In addition, we wanted the texts to be evenly distributed over the time span of the legislative term. To achieve this goal, we selected specific agenda items that covered a range of topics, and then sampled all speeches that belong to this specific agenda item, to increase the comparability of the contributions made by the different speakers.

We based our topic selection on the coding scheme developed in the Comparative Agendas Project (CAP) (Bevan, 2019). The CAP scheme includes 21 major topics and more than 200 fine-grained subtopics. We used a topic classifier to select speeches for eight of the major CAP topics for annotation (*Cultural Policy Issues, Defense, Domestic Macroeconomic Issues, Education, Environment, Health, Immigration and Refugee Issues, Law, Crime, Family Issues*) and manually validated the results.[6]

Annotation Our annotation follows a lexicographic approach, based on the automatically created German opinion verb lexicon of Wiegand and Ruppenhofer (2015). The lexicon includes 1,416 verbal subjective expressions, categorised as either *agent* (533), *patient* (141) or *speaker view* verbs (742). Our annotation setup proceeds as follows. We mark all verbs from the lexicon in

[6]For more detailed information, please refer to the data sheet in our github repository: https://github.com/umanlp/GePaDe-ORL.

our data for annotation and ask our annotators to disambiguate the view as either *agent, patient* or *speaker* view. If the verb can not be interpreted as a subjective expression in this particular context, then we assign the label *none*. After disambiguating the subjective expressions, the annotators are instructed to identify the holder and target for this subjective expression.[7]

In addition to *holder* and *target*, we annotate the *effect* role for patient and speaker view (see Ex. (1) below). We use the label *other* to encode a set of verb-specific roles (such as Cause, Theme, Goal) for speaker view verbs (see examples in Table 1).

(1) (Der Fall Susanna)$_{Target}$ zeigt beispielhaft (den Maximalschaden der Durchwinkekultur)$_{Effect}$.
(The Susanna case)$_{Target}$ shows (the maximum damage caused by the wave-through culture)$_{Effect}$.

The *particle* role (**PTC**) marks separated verb particles, as shown in Ex. (2) where the verb form "verlorengehen" (be lost) has a meaning very different from "gehen" (go) alone without the verb particle. To encode the actual meaning of the verb, we mark the separated verb particle as PTC. In addition, we use this label for obligatory reflexive pronouns (see Fig. 1).

(2) Über viele Jahrhunderte gewachsenes kulturelles Kapital geht hier (verloren)$_{Ptc}$.
Cultural capital that has grown over many centuries is being lost here.

We use the label **SVC** to indicate the nominal component of a support verb construction where the meaning is largely shifted from the verb to the noun, as illustrated in Ex. (3).

(3) Zeigen Sie endlich (Rückgrat)$_{SVC}$.
Finally show some (backbone)$_{SVC}$.

Our annotated dataset has a size of 214,229 tokens and 13,222 clauses.[8] The number of annotated subjective expressions and their roles is shown in Table 2. The numbers refer to SE and role counts where each role can consist of multiple tokens.

The annotation has been done independently by two trained student assistants. Throughout the annotation, we had weekly meetings to discuss open questions and difficult cases. After the coding has been completed, all disagreements have been resolved by a trained linguist and further consistency checks have been made to assure the quality of the data. We computed inter-annotator agreement (IAA) between the two students for role assignment as precision, recall and f-score on the token level. We first considered Annotator1 as the ground truth and evaluated Annotator2's predictions against A1. Then we switched roles and report the averaged agreement as prec: 74.83%, recall: 74.90%, and F1: 74.27%.

[7] While the lexicon specifies the view of each verb, some of the verbs also have other senses that belong to a different view and thus need to be disambiguated.

[8] We used spacy for sentence splitting which results in segments at the clause level, with an average size of around 16 tokens/clause.

	Agent	**Patient**	**Speaker**	**Total**
SE	2,325	138	859	3,322
Roles (all)	4,594	278	1,503	6,375
Target	2,422	109	752	3,283
Holder	1,998	116	12	2,126
Other	1	0	643	644
PTC	142	4	53	199
SVC	31	5	38	74
Effect	0	44	5	49

Table 2: Distribution of roles and views in our new data set. The numbers refer to counts on the **SE/role** level. PTC: separated verb prefixes and obligatory reflexive pronouns; SVC: support verb constructions.

Error analysis One frequent error concerns roles where one annotator assigned a specific label and the other coder also marked the same span but forgot to select a label for this span. Another frequent source of disagreements regards the selection of the role spans. Our student annotators had a background in political and social sciences and therefore sometimes struggled to identify the correct syntactic phrase for role annotation, as illustrated below. Here, A1 correctly chose the relative pronoun for target annotation while A2 assigned the target label to the head of the relative clause.

A1: die Frostschäden, (unter denen)$_{Target}$ (die Obstbauern)$_{Holder}$ zu leiden hatten
A2: (die Frostschäden)$_{Target}$, unter denen (die Obstbauern)$_{Holder}$ zu leiden hatten
Gloss: (the frost damage)$_{A2}$, (under which)$_{A1}$ (the fruit_growers)$_{Holder}$ to suffer had
Translation: *the frost damage suffered by fruit growers*

In a similar vein, we observed cases where one annotator had marked the whole noun phrase (as specified in the annotation guidelines) while A2 marked only the head of the noun phrase but left out modifier phrases or complement clauses attached to the head. This shows that for this type of annotation, linguistic training is more important than a background in political or social sciences.

5. Evaluation

We now present an evaluation where we assess how well an automatic system can predict the subjective expressions and opinion roles in our new dataset.

5.1. Experimental Setting

We split our data into training, development and test sets with 9,298/927/3,067 sentences, respectively. We ensure that none of the agenda items in the test set are included in the training set which results in a more challenging and realistic setting compared to distributing speeches from the same agenda item into training and test sets. This amounts to 177/18/72 (train/dev/test) different speeches, with 2,302 (train), 257 (dev) and 763 (test) annotated subjective expressions.

Baseline system The structure of our data is similar to semantic roles (see Fig. 1), which allows us to train a state of the art Semantic Role Labelling (SRL) system on our data. We chose the SRL system of Conia and Navigli (2020), a language- and syntax-agnostic model that jointly learns to predict the predicates, their senses and arguments (i.e., opinion roles). The model combines a predicate-aware word encoder with a predicate-argument encoder. The first component yields contextualised word representations with respect to the predicate of the sentence, while the second encoder learns predicate-aware argument representations. We initialise the model with the pretrained gbert-large[9] language model (Chan et al., 2020) and select the best fine-tuned model on the development set.[10]

Evaluation metric We report precision, recall and F1 (micro) for the prediction of subjective expressions and roles. Note that, due to our lexicographic approach, the position of all potential SEs are given (hence recall for SE prediction is 100%) and the system only has to decide whether a given verb form at position i is a subjective expression (SE) or not (none).[11] As the role labels can cover more than one token, they are therefore represented as sets of (possibly discontinuous) tokens. The annotation scheme assumes that a given verb can bear at most one SE annotation, that is, it can evoke at most one instance of subjective expression. For roles this is not true: a set of tokens could bear multiple role labels, usually in relation to different SEs. According to our annotation guidelines, roles are dependent on SEs and so system roles can match gold roles only if they are related to the same SE. In line with this, the evaluation first checks how system SEs and gold SEs align. System SEs that cannot be aligned to gold SEs produce false positives, including for their associated roles. In symmetric fashion, gold SEs that cannot be aligned to a system SE result in false negatives. For roles, alignment requires non-zero overlap with the tokens covered by a label of the same type on the other side. Each component token of aligned labels is counted as a true or false positive, or as a false negative. This means that longer spans contribute more to the overall score than shorter labels.

[9] https://huggingface.co/deepset/gbert-large

[10] To ensure replicability, we will release the configuration files together with the train/dev/test splits.

[11] The data includes 3,322 subjective expressions and 1,167 non-subjective uses of those verb forms, i.e., 26% of the candidate expressions have the label NONE.

	Prec	Rec	F1	Prec	Rec	F1
SE	93.0	100	96.3	93.0	100	96.4
Roles	74.0	74.8	74.4	70.7	76.8	73.6
Target	77.7	78.2	77.9	71.2	82.0	76.3
Holder	76.0	78.9	77.4	75.9	84.7	80.0
Other	57.7	59.6	58.5	69.1	55.9	61.6
PTC	41.1	54.5	46.2	68.9	56.8	60.0
SVC	33.7	14.6	20.3	20.5	14.3	16.7
Effect	42.0	47.1	44.4	22.5	85.0	35.5

Table 3: Precision, recall and F1 (micro) for SE prediction and roles (token overlap). Results are averaged over three runs with different initialisations (white: dev set, gray: test set).

Results Table 3 shows precision, recall and F1 (micro) for SEs and roles on the development (white) and test set (gray). Precision for the prediction of subjective expressions is around 93% for all runs, with a standard deviation of 0.33/0.68% on the dev/test set. This shows that the system has no problem to distinguish between subjective and non-subjective uses in our data.

Results for roles are substantially lower with around 73% micro-F1 for all roles. However, results for holders and targets, which are at the center of our interest, are still high with an F1 in the range of 77-80%. The gap in results between holders and targets can be explained by their length. Holders in our corpus have an average length of 1.5 tokens while targets are much longer with 5.5 tokens on average, making them more challenging to predict. For the other, less frequent roles, however, results are much lower as there are not enough annotations for the model to learn.

6. Conclusions

In the paper, we presented a new dataset for German political debates, with manual annotations for subjective expressions and their opinion roles. We showed that we can use an SRL system to identify subjective expressions and their holders and targets in text with good prediction accuracy. In future work, we plan to apply our system to predict opinion holders and their targets in a large corpus of parliamentary debates, to study the sources and targets of opinions for specific topics across speakers and parties.

7. Limitations

An important limitation of our work is that our corpus only includes annotations for one language (German) and text type (parliamentary debates). However, we expect that our approach can be easily extended to similar text types such as party press releases, manifestos or newspaper articles and plan to investigate this in future work. Another weakness of our work are the low results for the low-frequency labels. As we are mostly interested in the identification of holders and targets, this is not a severe problem but we strongly recommend users who apply our model not to rely on the predictions for the other labels.

Acknowledgements

The work presented in this paper was funded in part by a research grant from the Ministry of Science, Research and the Arts (MWK) Baden-Württemberg and by the German Research Foundation (DFG) under the UNCOVER project (RE3536/3-1). We would also like to thank our student annotators for their dedicated work.

8. Bibliographical References

Gavin Abercrombie and Riza Batista-Navarro. 2018. 'Aye' or 'No'? Speech-level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

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

Gavin Abercrombie and Riza Batista-Navarro. 2022. Policy-focused Stance Detection in Parliamentary Debate Speeches. *Northern European Journal of Language Technology*, 8(1).

Gavin Abercrombie and Riza Theresa Batista-Navarro. 2018. Identifying Opinion-Topics and Polarity of Parliamentary Debate Motions. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 280–285, Brussels, Belgium. Association for Computational Linguistics.

Gavin Abercrombie, Federico Nanni, Riza Batista-Navarro, and Simone Paolo Ponzetto. 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. Association for Computational Linguistics.

Valentin Barriere, Alexandra Balahur, and Brian Ravenet. 2022a. Debating Europe: A multilingual multi-target stance classification dataset of online debates. In *Proceedings of the LREC 2022 workshop on Natural Language Processing for Political Sciences*, pages 16–21, Marseille, France. European Language Resources Association.

Valentin Barriere, Guillaume Guillaume Jacquet, and Leo Hemamou. 2022b. CoFE: A new dataset of intra-multilingual multi-target stance classification from an online European participatory democracy platform. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 418–422, Online only. Association for Computational Linguistics.

Samuel E. Bestvater and Burt L. Monroe. 2023. Sentiment is not stance: Target-aware opinion classification for political text analysis. *Political Analysis*, 31(2):235–256.

Shaun Bevan. 2019. Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook. In Frank R. Baumgartner, Christian Breunig, and Emiliano Grossman, editors, *Comparative Policy Agendas: Theory, Tools, Data*. Oxford: Oxford University Press.

Branden Chan, Stefan Schweter, and Timo Möller. 2020. German's next language model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6788–6796, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Simone Conia and Roberto Navigli. 2020. Bridging the gap in multilingual semantic role labeling: a language-agnostic approach. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1396–1410, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Anne Göhring, Manfred Klenner, and Sophia Conrad. 2021. DeInStance: Creating and evaluating a German corpus for fine-grained inferred stance detection. In *Proceedings of the 17th Conference on Natural Language Processing*

(KONVENS 2021), pages 213–217, Düsseldorf, Germany. KONVENS 2021 Organizers.

Rincy Jose and Varghese S Chooralil. 2015. Prediction of election result by enhanced sentiment analysis on twitter data using word sense disambiguation. In *2015 International Conference on Control Communication & Computing India (ICCC)*, pages 638–641.

Allison Koh, Daniel Kai Sheng Boey, and Hannah Béchara. 2021. Predicting policy domains from party manifestos with BERT and convolutional neural networks. In *Proceedings of the 1st Workshop on Computational Linguistics for Political Text Analysis (CPSS-2021)*, pages 67–77, Düsseldorf, Germany.

Gabriella Lapesa, Andre Blessing, Nico Blokker, Erenay Dayanik, Sebastian Haunss, Jonas Kuhn, and Sebastian Padó. 2020. DEbateNet-mig15:tracing the 2015 immigration debate in Germany over time. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 919–927, Marseille, France. European Language Resources Association.

Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 02(97):311–331.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Isa Maks and Piek Vossen. 2011. A verb lexicon model for deep sentiment analysis and opinion mining applications. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 10–18, Portland, Oregon. Association for Computational Linguistics.

Laura Mascarell, Tatyana Ruzsics, Christian Schneebeli, Philippe Schlattner, Luca Campanella, Severin Klingler, and Cristina Kadar. 2021. Stance detection in German news articles. In *Proceedings of the Fourth Workshop on Fact Extraction and VERification (FEVER)*, pages 66–77, Dominican Republic. Association for Computational Linguistics.

Dhiraj Murthy. 2015. Twitter and elections: Are tweets, predictive, reactive, or a form of buzz? *Information, Communication & Society*, 18(7):816–831.

Sven-Oliver Proksch, Will Lowe, Jens Wäckerle, and Stuart Soroka. 2019. Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly*, 44(1):97–131.

Christian Rauh. 2018. Validating a sentiment dictionary for German political language—a workbench note. *Journal of Information Technology & Politics*, 15(4):319–343.

Rezvaneh Rezapour, Lufan Wang, Omid Abdar, and Jana Diesner. 2017. Identifying the Overlap between Election Result and Candidates' Ranking Based on Hashtag-Enhanced, Lexicon-Based Sentiment Analysis. In *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, pages 93–96, San Diego, CA, USA. IEEE.

Josef Ruppenhofer, Julia Maria Struš, Jonathan Sonntag, and Stefan Gindl. 2014. Iggsa-steps: Shared task on source and target extraction from political speeches. *Journal for Language Technology and Computational Linguistics*, 29(1):33 – 46.

Josef Ruppenhofer, Julia Maria Struss, and Michael Wiegand. 2016. Overview of the iggsa 2016 shared task on source and target extraction from political speeches. In *Proceedings of IGGSA Shared Task 2016 Workshop*, pages 1–9.

Jonathan B Slapin and Sven-Oliver Proksch. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52:705–722.

Veselin Stoyanov, Claire Cardie, Diane Litman, and Janyce Wiebe. 2004. Evaluating an opinion annotation scheme using a new multi-perspective question and answer corpus. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.

Shivashankar Subramanian, Trevor Cohn, Timothy Baldwin, and Julian Brooke. 2017. Joint sentence-document model for manifesto text analysis. In *Proceedings of the Australasian Language Technology Association Workshop, ALTA 2017, Brisbane, Australia, December 6-8, 2017*, pages 25–33.

Samson Ebenezar Uthirapathy and Domnic Sandanam. 2023. Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, 218:908–917.

Jannis Vamvas and Rico Sennrich. 2020. X-stance: A multilingual multi-target dataset for stance detection. In *5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*. CEUR Workshop Proceedings.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005a. Annotating expressions of opinions and emotions in language. *Lang. Resour. Evaluation*, 39(2-3):165–210.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005b. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39:165–210.

Michael Wiegand and Josef Ruppenhofer. 2015. Opinion holder and target extraction based on the induction of verbal categories. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 215–225, Beijing, China. Association for Computational Linguistics.