Who is *we*? Disambiguating the referents of first person plural pronouns in parliamentary debates

Ines Rehbein
Data and Web Science Group
University of Mannheim

Josef Ruppenhofer
Archive for Spoken German
Leibniz Institute
for the German Language

Julian Bernauer
MZES
University of Mannheim

ines@informatik.uni-mannheim.de   ruppenhofer@ids-mannheim.de
julian.bernauer@mzes.uni-mannheim.de

Abstract

This paper investigates the use of first person plural pronouns as a rhetorical device in political speeches. We present an annotation schema for disambiguating pronoun references and use our schema to create an annotated corpus of debates from the German Bundestag. We then use our corpus to learn to *automatically* resolve pronoun referents in parliamentary debates. We explore the use of data augmentation with weak supervision to further expand our corpus and report preliminary results.

1 Introduction

Personal pronouns are an important rhetorical device in political speeches that allow politicians to shape their message to appeal to specific audiences. Multiple functions of pronouns have been described, such as creating a feeling of unity with the audience (1.1), sharing responsibility (1.2) or criticising others (1.3) (Beard, 2000; Bramley, 2001; Håkansson, 2012).

Example 1.1. Members of Congress, we must work together to help control those costs (Bush 2004)

Example 1.2. We have increased our budget at a responsible 4 percent (Bush 2001)

Example 1.3. the more we get involved with other people, the more complicated our relationships get (B. Clinton 2002) [1]

Tyrkkö (2016) calls personal pronouns "one of the primary linguistic features used by political speakers to manage their audiences' perceptions of in-groups and out-groups". This makes them especially important for populist rhetoric where the speaker evokes a dichotomous view of society,

us-versus-them (see, e.g., Mudde (2004); Mudde and Kaltwasser (2017)).

While the practice of *othering* might seem to be the most prominent feature of personal pronouns in political discourse, another important aspect also needs to be considered, namely their referential ambiguity (Tyrkkö, 2016; Wales, 1996). As stated by Allen (2007, pp.12),

> "Shifting identity through pronoun choice and using pronouns with ambiguous referents enables politicians to appeal to diverse audiences which helps broaden their ability to persuade the audience to their point of view. It is a scattergun effect —shoot broadly enough and you'll hit something".

While prior research on the interface between corpus linguistics, pragmatics, discourse studies and political science has presented empirical findings based on word frequencies (Vuković, 2012; Tyrkkö, 2016; Alavidze, 2017), only few studies have tried to systematically investigate this topic in more detail, i.e., by trying to measure the agreement between human annotators for disambiguating the referents of personal pronouns in political speeches, or by presenting large-scale studies of the use of personal pronouns beyond word frequencies.

This paper takes first steps in that direction by means of an annotation study in which we classify instancse of the first person plural pronoun *wir* 'we' in German parliamentary debates, using a classification scheme with 9 different classes. We report inter-annotator agreement for this highly subjective task and analyse our disagreements. We then present a preliminary analysis of our data where we look into differences in the use of *we/us* in political speeches, depending on (i)

---

[1] Two of the examples taken from Håkansson (2012).

the speaker, (ii) the topic, and (iii) the speaker's party affiliation.

In the second part of the paper, we undertake first experiments towards automatically predicting the referents of first person pronouns in parliamentary debates. For that, we make use of transfer learning, in combination with data augmentation based on weak supervision (Ratner et al., 2016, 2020). We show that our transfer learning approach brings substantial improvements over a majority baseline while pretraining the model on the larger, noisy data and fine-tuning it on our manual annotations yields only small improvements over training on the manual annotations only.

2 Related Work

First person plural pronouns from a linguistic perspective The reference of German *wir*, just like that of English *we*, is quite variable. Following the typology of Cysouw (2002), German *wir* as a first person plural (1PL) form has multiple distinct uses: (i) `minimal inclusive`, consisting of speaker and hearer (2.1); (ii) `augmented inclusive`, adding third parties beyond the minimal inclusive (2.2); (iii) `exclusive`, consisting of the speaker and third parties, but excluding the hearer (2.3).

Example 2.1. Sollen wir morgen telefonieren?
'Shall we talk on the phone tomorrow?'

Example 2.2. Kim kommt um 12 an. Sollen wir dann Mittag essen gehen?
'Kim will arrive at 11. Shall we go to lunch then?' [all three of us]

Example 2.3. Wir gehen ins Kino. Was habt ihr vor?
'We're going to the movies. What are your plans?'

In addition, special subtypes of uses may be recognized. For English, Quirk et al. (1985) discuss a set of special (sub)uses that also occur with German *wir*. For instance, a single author may nevertheless use 1PL pronouns to avoid appearing 'egotistical'. Doctors (among others) may use the 1PL pronoun in a a hearer-oriented way (e.g. *How are we feeling today?*). Of greatest relevance to our data are Quirk et al. (1985)'s generic uses and their class of rhetorical uses where the pronoun refers to a collective such as 'the party', 'the nation'.

While linguistic analyses of pronouns often simply view them as words with determinate ref-

erence to a deictically, anaphorically or cataphorically available entity, pragmatic and discourse-oriented studies of pronouns like ours focus on their conceptual emptiness and the fact that their referents must be inferred in context, with the possibility of (un)intentionally ambiguous uses, since individuals have multiple social, discursive and interactional roles.

Corpus studies of 1PL reference Very close in spirit to our work but operating on conversational interactions and with categories appropriate to that domain, Scheibman (2014) presents a study on the reference of *we* in relation to predicate patterns and pragmatic functions. The study coded instances of *we* from the Santa Barbara Corpus of Spoken American English for several features, among them (i) the inclusive vs exclusive distinction, (ii) type of referent (e.g. family member, couple, classmates, human beings, etc.), (iii) tense of predicate, (iv) modals present. The authors' findings suggest that different referential uses of first personal pronouns may be distinguishable based on contextual cues such as tense and modality.

Pronouns in political discourse Tyrkkö (2016) presents a diachronic study of the use of personal pronouns in political speeches over two centuries, showing shifts from a self-centric style (marked by frequent use of *I*) towards the more inclusive use of 1PL forms in the 1920s, which the author ties to the emergence of broadcast media. The study does not disambiguate 1PL forms but counts all of them as inclusive.

Íñigo-Mora (2004) studies the use of *we* in 5 Question Time Sessions of the British parliament, where MPs ask questions of government ministers. She distinguishes what she calls exclusive, inclusive, generic and parliamentary uses of *we* and examines their distribution across different combinations of interactants (opposition MP to member of government; member of government to opposition MP; member of government and supportive MP (in either direction)).[2] The frequency distribution is interpreted along two dimensions: (i) power and distance and (ii) identity, community and persuasion. Among the findings

---

[2]There is no generally agreed-upon terminology used to distinguish uses of *we*, either in general or in the political or parliamentary context. For Inigo-Mora the generic *we* refers to "a kind of patriotic "we" that embraces all British people". In the terminology of Quirk et al. (1985) this would be called a collective use. In our annotation scheme, the uses at issue would be labeled "COUNTRY".

is that exclusive uses of *we* constitue the most common type overall, accounting for 53.4% of all tokens. Exclusive *we* is at its most dominant in interactions from government supporting MPs to opposition MPs (76.1%) while it is hardly ever used in questions from opposition MPs to a member of government, which is taken to reflect the power dynamics. Inclusive uses of *we* were found to be much rarer overall, making up 14.5% of all tokens. None of these are uttered by opposition members speaking to members of government, while three quarters are produced between government supporting MPs and members of government, expressing shared identity. Opposition MPs mostly use generic and parliamentary *we*, thus affiliating themselves with the parliament as a distinct branch of government and the country at large, likely because that is where persuasion is most likely to succeed. It is unclear to what extent these results carry over to the plenary setting.

Non-parliamentary political discourse   Studies of 1PL pronouns have also targeted other types of interactions. Bull and Fetzer (2006) analyze the use of *you* and *we* in tv interviews with British politicians that were broadcast during the 1997 and 2001 British general elections and just before the war with Iraq in 2003. The focus of the study was on question-response sequences in which politicians make use of pronominal shifts as a means of equivocation to effect shifts of accountability and responsibility. Proctor and Su (2011) examine the use of *we* by four (vice-)presidential candidates in debates and interviews around the time of the 2008 US election. The study focuses on which groups are the referents of *we* and which entities are picked out by possessive NPs of the form *our N*, considering the results in light of the candidates' political stature and targeted office as well as the differences between debate and interview settings.

Politeness   Finally, we note that quite a lot of research on pronoun use exists in the area of politeness, though this typically targets pronouns of address. For instance, in a seminal study, Brown and Gilman (1960) discussed the differences in use between informal and formal second person pronouns (such as German *du* and *Sie*) as forms of address in terms of their association with the dimensions of power and solidarity between speakers. The authors argue that, while for a long

Party	#Tokens	#Annot	#Spk	per 1000
AfD	8,993	142	8	15.8
CDU/CSU	10,674	335	5	31.4
FDP	7,358	166	7	22.6
GRÜNE	7,457	136	5	18.2
LINKE	9,310	130	6	14.0
SPD	7,438	245	4	32.9
fraktionslos	797	9	1	11.3
Total	52,027	1,163	36	22.3

Table 1: Some statistics for the annotated testset (**#Spk**: no. of speakers per party; **per 1000**: no. of 1PL pronouns per 1000 tokens).

time the form chosen was mainly determined by power differentials, over time the choice came to depend more on the factor of solidarity.

3   Annotation Study

3.1   Data

The data we use in our study are parliamentary debates from the German Bundestag, covering a time period from Oct 24, 2017 to May 19, 2021.[3] The corpus includes over 330,000 sentences (>16,5 mio tokens), with political speeches by 777 different speakers.

From the XML files, we extracted the individual speeches and randomly selected a subset for manual annotation where we tried to collect roughly the same number of speeches/tokens for each party (see table 1). This resulted in a testset with 36 speeches by different speakers (52,027 tokens) where we manually disambiguated all instances of first person plural pronouns (*wir, uns, unser, unsere, unseren, unseres, unsre*) by classifying them into nine predefined classes. We describe our annotation schema below (§3.2).

3.2   Annotation schema

Table 2 and Table 10 in the appendix give an overview over our classification schema. We assume that references of *we/us* in parliamentary debates can be assigned to a small number of different categories, such as "we, the PARLIAMENT" or "our COUNTRY", or "our political PARTY". The schema has been designed in a bottom-up, data-driven fashion, using speeches from the European parliament and the German Bundestag for schema development. We test our classification schema in an annotation experiment and investigate a) how well human annotators agree when

---

[3] The data is available in XML format from `https://www.bundestag.de/services/opendata`.

Class	Description	Example
BOARD	Members of a board/ commission/committee	**Wir** haben heute im Untersuchungsausschuss erfahren
COUNTRY	references to Germany/ all Germans	**Wir** sind Weltmeister **Unser** Grundgesetz
GENERIC	generic uses that can be replaced by *one (de: man)*	Daran werden **wir uns** noch in 100 Jahren erinnern
GOVERN	members of the government	**Wir** haben die Arbeitslosigkeit bekämpft.
PARL	members of the parliament	**Wir** Abgeordnete... Lassen Sie **uns** diesen Antrag heute beschließen
PARTY	members of one specific party	**Wir** Liberale haben schon früher...
PEOPLE	groups of people defined by social variables (age, profession, religion and other shared characteristics ...)	Wie **wir** Älteren **uns** verhalten... **Wir** Steuerzahler, **Wir** Christen, **Wir** Pendler, ...
SPECPERS	groups of individuals or members of more than one group	**Wir** beide haben darüber diskutiert **Wir**, die deutsche und die israelische Regierung
UNION	geo-political groups on a supranational level (EU, NATO)	**Wir** in der EU... **Unsere** Europäische Union...

Table 2: Overview of the annotation scheme for 1PL references in parliamentary debates.

disambiguating 1PL pronouns in political speech; b) whether it is possible to automatically predict the intended reference of personal pronouns in parliamentary debates.

We expect that, as noted in section 2, a large part of vagueness and ambiguity in political speech is intended and will result in low IAA between some of the classes in our classification schema. However, we also expect that some classes (such as PARTY) are less ambiguous which should be reflected in a higher agreement between the annotators.

3.3 Annotation

The annotators, two computational linguists,[4] were presented with the speech texts where all instances of 1PL pronouns were highlighted. The task then consisted in assigning a label to each pronoun.[5] The annotators were only allowed to assign exactly one label per instance.

Inter-Annotator Agreement (IAA) We report Krippendorff's α and percentage agreement for two annotators on the 1,163 annotated instances. Inter-rater agreement was quite high with 0.82 α. Table 3, however, shows substantial differences in agreement between the individual classes. We obtained very high agreement for COUNTRY and PARTY (> 90% F1) and slightly lower but still reasonably high agreement for GOVERNMENT, PARLAMENT and UNION (between $78-87\%$ F1). For GENERIC, PEOPLE and SPECIFIC_PERSONS, agreement was substantially lower ($58-66\%$ F1). Those

---

[4]The data was annotated by the first two authors of the paper.

[5]We used INCEpTION (Klie et al., 2018) as annotaton tool.

Class	F1	Support
BOARD	0.0	1
COUNTRY	92.0	411
GENERIC	65.2	67
GOVERN	87.2	167
PARL	86.6	299
PARTY	90.6	103
PEOPLE	66.7	13
SPECPER	58.8	20
UNION	78.2	82
Total	86.1	1,163

Table 3: IAA (F1) and support (number of annotated instances in the gold standard) for individual classes.

classes are also less frequent in the data. The remaining class, BOARD, was too rare in our testset to report meaningful results (1 instance only).[6] We kept this class despite its low frequency in the Bundestag corpus, as we found it to be more frequent in speeches from the European Parliament.

After the annotation was completed, the two annotators discussed and resolved all disagreements to create a ground truth dataset that we used as evaluation data in our experiments (§6).

4 Data Analysis

We now present a preliminary analysis on our manually annotated dataset where we focus on differences in the use of 1PL pronouns across politicians and parties.

Table 1 shows that the governmental parties produce the most 1PL instances per 1000 words, which makes sense given that their members can choose between the greatest number of collective

---

[6]The confusion matrix for the annotations can be found in the appendix, Table 11.

Party	BOARD	COUNTRY	GENERIC	GOVERN	PARL	PARTY	PEOPLE	SPECP	UNION
AfD	0.0 (0)	6.0 (54)	0.6 (5)	0.0 (0)	5.1 (46)	3.4 (31)	0.4 (4)	0.2 (2)	0.0 (0)
CDU/CSU	0.0 (0)	11.4 (122)	2.1 (22)	9.6 (102)	5.0 (53)	0.5 (5)	0.3 (3)	0.4 (4)	2.2 (24)
FDP	0.0 (0)	5.7 (42)	1.6 (12)	0.0 (0)	6.1 (45)	5.2 (38)	0.0 (0)	0.5 (4)	3.4 (25)
GRÜNE	0.0 (0)	5.9 (44)	1.7 (13)	0.1 (1)	7.8 (58)	1.2 (9)	0.5 (4)	0.7 (5)	0.3 (2)
LINKE	0.1 (1)	7.1 (66)	0.9 (8)	0.0 (0)	3.7 (34)	1.7 (16)	0.2 (2)	0.0 (0)	0.3 (3)
SPD	0.0 (0)	10.6 (79)	0.9 (7)	8.6 (64)	8.1 (60)	0.5 (4)	0.0 (0)	0.4 (3)	3.8 (28)
frakt.los	0.0 (0)	5.0 (4)	0.0 (0)	0.0 (0)	3.8 (3)	0.0 (0)	0.0 (0)	2.5 (2)	0.0 (0)
Total	1	411	67	167	299	103	13	20	82

Table 4: Distribution of classes in the annotated testset (frequency per 1000 tokens and raw counts in brackets).
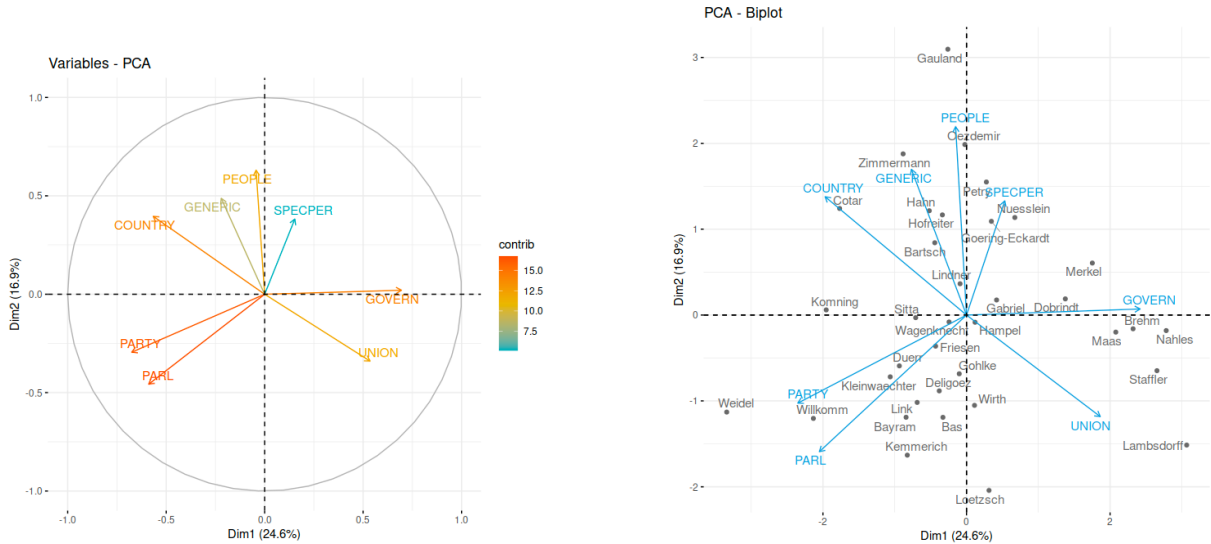

Figure 1: Principal Components Analysis (PCA): left figure shows the loadings for our class variables along the first two components (PC1, PC2), right figure also plots the speakers for PC1 and PC2.

identities.

Table 4 shows the distribution of the different classes across parties. As expected, only members of the CDU/CSU and SPD, the two parties involved in the government at the time of data collection, used *we* to refer to the government. Notably government MPs invoke their GOVERN identity substantially more than their PARTY identity. By contrast, members of the opposition parties refer more often to their own party, often to criticise the government and to distinguish their own policies from those of the government. This is particularly true for the FDP and the AfD, and to a lesser extent also for the LINKE and the GRÜNE.

All parties make frequent references to the parliament (PARL). The two parties in government, however, use many more references to COUNTRY than the opposition parties. This observation is in contrast to the findings of Íñigo-Mora (2004) (see Section 2) who found more pronoun references to the country from members of the opposition. We would like to stress that our data is not yet

large enough to produce representative results. In addition, we would also expect an impact of interaction type on the use of pronouns. Íñigo-Mora (2004) investigated Question Time sessions in the British parliament while we focus on plenary speeches, which are longer, less interactive and always have a mixed audience of supporters and opponents, whereas Question Time (superficially) addresses only one or the other. These differences might be reflected in different communicative strategies and stylistic choices.

Another reason for the higher ratio of COUNTRY references in speeches by members of the governmental parties may be that their ranks include key office holders such as the minister of foreign affairs, whose topics tend to skew (inter)national. To investigate this, more data is needed so that we can control for the effects of office holders.

Figure 1 (left) shows the loadings for our class variables along the first two dimensions of a Principal Components Analysis (PCA), based on the normalised frequency counts for the different

class variables for individual speakers. The first dimension (X axis) reflects 1PL pronoun references to the government on the right-hand side and to specific parties or the parliament as a whole on the left-hand side. This opposition separates politicians from the governmental parties from the ones from the opposition parties along the first dimension (Figure 1, right).

Figure 1 (right) also seems to show topical effects as Lambsdorff, a member of the FDP and the EU parliament, is positioned closest to the vector showing the loadings of the UNION variable. This might explain why he, as the only non-governmental politician, is also positioned at the right end of the first dimension. The politicians that are positioned left-most on the first dimension are Weidel (AfD), Willkomm (FDP), Komning (AfD) and Cotar (AfD). For the members of AfD, a nationalist and right-wing party deeply opposed to the European Union, it seems plausible that they are positioned not only at the opposite end of GOVERN but also of UNION. Further analysis is needed to investigate this.

Figure 1 (left) also shows that while the two classes PARTY and PARL are highly correlated and in opposition to GOVERNMENT, the more generic classes COUNTRY, GENERIC and PEOPLE also seem to cluster together. This again seems like a promising start for a more detailed analysis. Once more data has been annotated, it will be interesting to include the topic of the speeches in the analysis. This can be easily done, either based on the agenda of the debates or by using topic models. At the moment, however, our data is still too sparse for a more fine-grained analysis.

5 Training Data Augmentation

We now investigate whether and how well we are able to resolve ambiguities in 1PL pronoun references in parliamentary debates *automatically*, using our small annotated dataset to train a supervised ML system.

As our manually annotated dataset is too small to expect high accuracies for automatic prediction, we resort to data augmentation with weak supervision. Our approach proceeds as follows.

We first extract text segments from parliamentary debates from the German Bundestag (19th legislative term) and remove the debates in the test set from our unlabelled training corpus. Each segment consists of a paragraph with multiple

Class	#Pattern	#Hits	err/N
BOARD	1	7	0/7
COUNTRY	5	8,795	2/25
GENERIC	4	307	8/25
GOVERN	5	14,851	2/25
PARLIAMENT	4	3,339	2/25
PARTY	11	8,265	4/25
PEOPLE	1	230	19/25
SPECPER	4	106	3/25
UNION	4	540	3/25
TOTAL	40	36,433	43/203

Table 5: Distribution of distinct patterns per class used for training data creation and number of hits for each pattern. Last column shows no. of errors in N randomly sampled pattern instances.

sentences, as annotated in the xml files. Please note that we do not assign labels to segments but to *instances of 1PL pronouns* in the segments. We then apply a set of predefined patterns to identify instances of 1PL pronouns for each class in our annotation scheme. With the help of these patterns, we assign labels to the unlabelled training corpus and can now use this data to train a supervised ML system for pronoun disambiguation. Below we explain the different steps in more detail.

Patterns For pattern extraction, we make use of the spaCy DependencyMatcher which provides a flexible and efficient framework for defining search patterns over dependency trees.[7]

We combine the spacy DependencyMatcher with the Snorkel framework (Ratner et al., 2016, 2020), a programmatic approach to data augmentation without manual labelling effort. Instead, Snorkel provides an API that allows users to write labelling functions that target specific labels in the annotation scheme. Those functions can consist of simple string matches but can also include more sophisticated features by including the predictions of pretrained classifiers or information from external knowledge bases. While these labelling functions are expected to have low coverage and might also introduce a certain amount of noise, Snorkel addresses this problem by learning an unsupervised generative model over the output of the labelling functions, based on the (dis-)agreements between the predicted labels. This approach is similar in spirit to previous work on

---

[7]See https://spacy.io/api/dependencymatcher. To generate the trees, we use the German de_core_news_sm model also provided by spaCy.

quality estimation for annotations obtained from crowdsourcing (Hovy et al., 2013). The output of Snorkel is a set of probabilistic labels that can be used as input to any supervised ML classifier.

Table 5 shows the number of patterns used for each class and the number of hits, i.e., instances extracted by each pattern from the unlabelled training data. Please note that the number of patterns is not very informative on its own, as patterns can make use of regular expressions, lemma lists and syntactic patterns over dependency trees, thus allowing us to extract a larger variety of diverse training examples than could be obtained based on simple string matches.

As an example, consider the following patterns used to extract labelled data for the PARTY class. Our first pattern looks for instances of *wir, uns* (we, us) directly followed by a party name. This pattern can extract instances like *Wir Grüne* or *uns Liberale*. Another pattern looks for instances of *wir* as the subject of communication verbs like *kritisieren, hinterfragen* (criticize, question) etc., as those are usually statements refering to specific parties from the opposition. A third example relies on future forms of *werden* (will) in combination with verbs of action, such as *schaffen, durchführen, investieren* (accomplish, execute, invest) to detect instances from the GOVERNMENT class. This pattern would extract matches like *wir werden Arbeitsplätze schaffen* 'we will create jobs' or *Mindestens 2 Mrd. EUR werden wir in den sozialen Wohnungsbau investieren* 'We will invest at least EUR 2 billion in social housing construction'.

The result of our pattern-based approach is a silver standard corpus with more than 36,000 labelled instances. To get an impression of the quality of the patterns, we randomly extracted 25 instances per class and manually inspected them (last two columns in Table 5). While most patterns seem to produce only a small amount of noise, some categories were more problematic. We found it particularly difficult to produce reliable patterns for PEOPLE and GENERIC which is reflected in the low coverage and precision for the two classes (see §6, Table 9).

6 Experiments

We now explore the potential of our automatically created training set for disambiguating references of personal pronouns in political debates.

wform	class	support	DL
wir	PARL	(185/600)	9
unser	COUNTRY	(24/26)	2
Wir	COUNTRY	(65/240)	9
unserem	COUNTRY	(28/32)	4
uns	COUNTRY	(56/163)	8
unsere	COUNTRY	(25/42)	6
unserer	COUNTRY	(19/31)	7
unseren	COUNTRY	(7/11)	4
Uns	PARL	(1/2)	2
Unser	COUNTRY	(4/5)	2
Unsere	COUNTRY	(3/4)	2
unseres	COUNTRY	(6/6)	1
unsre	COUNTRY	(1/1)	1
Unsre	COUNTRY	(2/2)	1
Total		(426/1163) Acc=36.6%	

Table 6: Majority baseline, support and no. of distinct labels (DL) per pronoun word form in the test set.

For that, we report results for three baselines and then present transfer learning experiments where we use our automatically created dataset for pre-training and then fine-tune the model on the manually created dataset.

B1: Majority Baseline Our first baseline assigns each pronoun word form its most frequent label (Table 6). This results in an accuracy of 36.6%. The last column shows the number of distinct labels (DL) per pronoun word form in the test set. The three most frequent word forms can occur with nearly any class (*Wir, wir*: 9 DL, *uns*: 8 DL), thus showing the difficulty of this task.

B2: Rule-based Baseline Our second baseline is a rule-based system that simply applies our pre-defined patterns to the testset and labels all matches with the respective labels. We use Snorkel's generative model (see §5) for resolving ties between conflicting rules and report precision, recall and F1 for the rule-based approach. Table 8 (B2) shows that while we obtain a reasonable precision for some patterns (COUNTRY: 92%, PARL: 91%, PARTY: 72%), recall is a huge problem. For the two most difficult patterns, GENERIC and PEOPLE, we obtain not even one correct match.

B3: Feature-based Classification Our third baseline makes use of a conventional feature-based approach to text classification. For that, we consider the following features: (1) tf-idf ngram features (unigrams, bigrams, trigrams) for the left and right context of each 1PL pronoun, (2) the word form of the pronoun, and (3) named entities in the left and right context of the pronoun. We explored different settings for these features

setting	value
left/right context size	20 tokens
bow unigrams	yes
bow bigrams	yes
bow trigrams	no
tfidf	yes
lemmatisation	yes
stopwords	no
feature selection	yes (χ^2)
num features	300
NER in left/right context	no

Table 7: Feature settings used for B3 (feature-based classification, Table 7).

in a 5-fold cross-validation setup and observed best results for the feature values show in Table 7. We tested different classifiers (linear SVM, Ridge regression, SGD, decision trees, AdaBoost, Random Forests) and found that linear SVM gave us best results on our data (49.3% acc.).[8] Table 8 (B3) shows results for the linear SVM classifier. Results for other models and settings were in the range of 35-47% acc.

Transfer Learning Model Our model uses a simple transformer architecture, based on the sentence pair classifier implementation of Simpletransformers[9] and the pretrained `bert-base-german-dbmdz-cased` model.[10] For details on parameter settings, please refer to Table 12 in the appendix. The motivation behind modelling personal pronoun disambiguation as sentence pair classification is that we want to make the model aware of the pronoun's left and right context. For that, we split each instance into two sequences where the first sequence encodes the left context of the pronoun in question and the second sequence includes the pronoun and its right context (see figure 2 below). Please note that our instances encode paragraphs, not sentences, and that S1 and S2 can thus include more than one sentences. In cases where the 1PL pronoun is positioned at the beginning of the paragraph, S1 will be empty.

Members of Congress ,	**we** must work ...
S1	S2

Figure 2: Setup for transfer learning using sentence pair classification; S1 encodes the left context of the 1PL pronoun, S2 the pronoun and its right context.

Results for 5-fold cross-validation We now report cross-validation results on our small, manually annotated dataset (Table 9). As we do not have enough data to create a representative validation set for model selection, we report preliminary results for all models (T1, T2, T3) after 25 epochs of training. This procedure has to be taken with a grain of salt and will be addressed, once we have more annotated data.

The results show that even a small number of annotated instances yields substantial improvements over the majority baseline (Table 6) and accuracy increases from 36.6% to over 50%. The results, however, are only slightly higher than the ones for the SVM (Table 8, B2). Table 9, (T2) shows results for merging the hand-annotated data with the noisy labels. In order not to outweigh the manual annotations, we downsampled the additional training data to at most 300 new instances per class. This setting results in only minor improvements (from 50.2 to 50.9% acc.). In our third setting, we use the noisy labels for an additional pretraining step before fine-tuning the model on the hand-annotated data. This yields another small improvement and increases accuracy to 51.8%.

Discussion The somewhat disappointing results for our data augmentation strategy might have several reasons. First, it is conceivable that we need to put more effort into creating a) more precise and b) more diverse rules, and c) to improve coverage. Results on a held-out dataset, created by the same rule-based approach, show that our model is perfectly able to learn the annotations in the weakly supervised data, achieving an accuracy of 97.6% on the held-out data. This shows that despite our efforts to minimise lexical cues and rely more on syntactic patterns, our augmented training data is highly biased and does not enable the model to learn good generalisations for each class.

While improving coverage for the rule-based approach might ameliorate the problem, it is also possible that the pattern-based approach is more

| | | **B2** | | | | | **B3** | | |
Class	#Gold	#Hits	TP	Prec	Rec	F1	Prec	Rec	F1
BOARD	1	0	0	0	0	0	0	0	0
COUNTRY	411	37	34	92	8	15	53	72	61
GENERIC	67	0	0	0	0	0	35	10	16
GOVERNMENT	167	53	23	45	15	22	41	35	38
PARLIAMENT	299	11	10	91	3	7	47	56	51
PARTY	103	17	13	76	13	22	49	30	37
PEOPLE	13	2	0	0	0	0	0	0	0
SPEC_PERSON	20	1	1	0	6	11	0	0	0
UNION	82	2	1	50	1	3	45	16	23
Total	1,163	123	83	Acc = 7.0%			Acc = 49.3%		

Table 8: Results for rule-based baseline (B2) and for the feature-based classification baseline (B3) (precision, recall and f1 for individual classes and acc. for all instances).

| | | **T1** | | | **T2** | | | **T3** | | |
Class	#Gold	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
BOARD	1	0	0	0	0	0	0	0	0	0
COUNTRY	411	58	65	62	65	63	64	56	66	60
GENERIC	67	29	13	18	20	16	18	50	7	13
GOVERNMENT	167	40	36	38	40	47	43	40	4	7
PARLIAMENT	299	50	64	56	56	54	55	45	78	57
PARTY	103	56	36	44	52	54	53	43	32	36
PEOPLE	13	0	0	0	9	23	13	0	0	0
SPEC_PERSON	20	17	10	12	6	1	8	0	0	0
UNION	82	28	17	21	36	24	29	60	9	16
Total	1,163	Acc = 50.2%			Acc = 50.9%			Acc = 51.8%		

Table 9: Results for 5-fold cross-validation for 3 transfer learning settings. T1: training on testset only; T2: training on testset + augmented data; T3: pretraining on augmented data and fine-tuning on testset (precision, recall and f1 for individual classes and acc. for all instances).

suitable for less ambiguous classification tasks, such as spam detection or offensive language detection, where we only have a small number of classes that are more clearly divided and where it is easier to create patterns with a high precision and coverage.

7 Conclusions

In the paper, we investigated what kinds of collectives 1PL pronouns refer to in parliamentary debates. To this end, we developed an annotation scheme that assigned references to one of nine categories and explored how well human annotators agree when assigning those categories. Our annotation study showed a substantial agreement of $> 0.8\alpha$ between two human raters. We then presented a preliminary analysis of the use of 1PL pronouns as a rhetorical device and pointed to some crucial differences between the parties as well as between members of the government and opposition parties. We subsequently explored how well we are able to automatically resolve ambiguous 1PL pronouns in parliamentary debates, using transfer learning and data augmentation. While our preliminary results are promising, there is room for improvement before we can apply our work to large-scale analysis of pronoun references in political text.

In future work, we plan to improve the accuracy of 1PL pronoun resolution by creating more training data, but also by improving the model itself. Possible ways to do so include providing the model with more information on the speaker, such as the speaker's name, party affiliation or whether or not the speaker is part of the government. Other improvements might come from jointly modelling 1PL pronouns in context, instead of looking at them one at a time.

Acknowledgments

References

Maia Alavidze. 2017. The use of pronouns in political discourse. *International Journal of Arts & Sciences*,

9(4):349–356.

Wendy Allen. 2007. Australian political discourse: Pronominal choice in campaign speeches. In *Selected Papers from the 2006 Conference of the Australian Linguistic Society, ed. by Mushin Ilana and Mary Laughren.*

Adrian Beard. 2000. *The Language of Politics.* London: Routledge.

Nicolette Ruth Bramley. 2001. *Pronouns of politics : the use of pronouns in the construction of 'self' and 'other' in political interviews.* Ph.D. thesis, Faculty of Arts and The Australian National University.

R. Brown and A. Gilman. 1960. The pronouns of power and solidarity. In T. A. Sebeok, editor, *Style in Language*, pages 253–276. MIT Press, Cambridge, Mass.

Peter Bull and Anita Fetzer. 2006. Who are we and who are you? the strategic use of forms of address in political interviews. *Text and Talk*, 26.

Michael Cysouw. 2002. The impact of an inclusive/exclusive opposition on the paradigmatic structure of person marking. In *Pronouns: Grammar and Representation*, pages 41–62. John Benjamins Publishing.

Jessica Håkansson. 2012. The use of personal pronouns in political speeches : A comparative study of the pronominal choices of two american presidents. School of Language and Literature, Linneaus University, Sweden.

Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani, and Eduard H. Hovy. 2013. Learning whom to trust with MACE. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 1120–1130. The Association for Computational Linguistics.

Isabel Íñigo-Mora. 2004. On the use of the personal pronoun we in communities. *Journal of Language and Politics*, 3(1):27–52.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics.

Cas Mudde. 2004. The populist zeitgeist. *Government and Opposition*, 39:541–563.

Cas Mudde and Cristóbal Rovira Kaltwasser. 2017. *Populism: A very short introduction.* Oxford, UK: Oxford University Press.

Katarzyna Proctor and Lily I-Wen Su. 2011. The 1st person plural in political discourse – American politicians in interviews and in a debate. *Journal of Pragmatics*, 43(13):3251–3266.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language.* Longman, London.

Alexander Ratner, Stephen H. Bach, Henry R. Ehrenberg, Jason A. Fries, Sen Wu, and Christopher Ré. 2020. Snorkel: rapid training data creation with weak supervision. *VLDB J.*, 29(2-3):709–730.

Alexander J. Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. 2016. Data programming: Creating large training sets, quickly. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3567–3575.

Joanne Scheibman. 2014. Referentiality, predicate patterns, and functions of we-utterances in American English interactions. In Theodossia-Soula Pavlidou, editor, *Constructing Collectivity: 'We' Across Languages and Contexts*, pages 23–43.

Jukka Tyrkkö. 2016. Looking for rhetorical thresholds: Pronoun frequencies in political speeches. *Studies in Variation, Contacts and Change in English*, 17.

Milica Vuković. 2012. Positioning in pre-prepared and spontaneous parliamentary discourse: Choice of person in the parliament of montenegro. *Discourse & Society*, (2):184–202.

Kate Wales. 1996. *Personal Pronouns in Present-Day English.* Cambridge: CUP.

A Appendices

Category	Description	Examples
COUNTRY	Refers to the country as a geo-political unit or to all citizens of this country. TEST: can be replaced by • "we Germans" • "our country" • "the German X"	**Wir** haben 2 Weltkriege verloren. **Wir** sind Exportweltmeister. **Wir** sind Papst. **Wir** als nationale Schicksalsgemeinschaft. **Wir** dürfen uns nicht vom Rest der Welt abschotten. **Unser** Grundgesetz / unsere Demokratie
PEOPLE	Refers to a (possibly large) group of people that are not defined by their nationality but by shared social variables or characteristics such as age, gender, class, religion, profession, ... Also used for references to society that are not limited to Germany as a geo-political unit.	wie **wir** Christen uns verhalten **Wir** als arbeitende Bevölkerung **Wir** Älteren, **Wir** Rentner **Wir** Steuerzahler, **Wir** Pendler
PARTY	Refers to members of a specific party (including coalitions of like-minded parties, e.g., on the supranational level)	**Unser** Antrag geht einen entscheidenden Schritt... **Wir** werden diese Regierung weiter kritisieren. **Wir** Liberale haben schon vor Jahren gesagt, ...
PARL	Refers to all members of the parliament (also references to both, government and opposition)	**Wir** Abgeordnete sind vom Volk gewählt. In diesem Haus debattieren **wir** heute... Lassen Sie **uns** diesen Antrag heute beschließen.
GOVERN	Refers to all members of the government	**Wir** haben entscheidende Schritte getan, um die Digitalisierung zu fördern. **Wir** haben Familien entlastet und die Arbeitslosigkeit bekämpft
UNION	Refers to geo-political groups on a supranational level, e.g., the EU, the NATO, etc.	**Wir** in der EU müssen zusammen einen Weg finden, wie wir unsere Sicherheitspolitik gestalten.
SPEC_PERS (GROUPS)	Refers to groups of specific individuals or members of more than one group	Sie haben die PKK und die YPG in einen Topf geworfen, **wir** sind aber nicht deckungsgleich. Frau Merkel und ich, **wir** haben darüber lange diskutiert. **Wir**, die deutsche und die israelische Regierung
GENERIC	Generic uses of *we/us* that can be replaced by *one/you* (German: man/es gibt) or *unser/e* can be replaced by *diese*. We assume a generic reading if *we/us* refers to the whole world/universe.	Das brauchen **wir** überall in der Welt → das braucht man überall... In den letzten Jahren haben **wir** viel über den Wandel der Gesellschaft gehört → hat man viel gehört über... Woran **wir** uns noch in 100 Jahren erinnern werden → Woran man sich noch in... die schwierigen Probleme **unserer** Zeit → dieser Zeit In einer Welt, in der **wir** über 222 gewaltsam ausgetragene Konflikte haben → in der es ... gibt
BOARD	Refers to members of a board / commission / committee /political organisation on the subnational level (subgroups of the parliament/government)	**Wir** haben im Untersuchungsausschuss viel diskutiert... Im Coronakabinett haben **wir** beschlossen... Im Agrarausschuss haben **wir** ...

Table 10: Overview of the annotation scheme for 1PL references in parliamentary debates.

A2 \ A1	BOARD	COUNTRY	GENERIC	GOVREN	PARL	PARTY	PEOPLE	SPECPER	UNION
BOARD	0	0	0	0	0	0	0	0	0
COUNTRY	0	385	8	4	14	3	1	3	12
GENERIC	0	4	46	1	13	0	2	0	1
GOVERN	0	7	1	146	8	7	0	1	4
PARL	1	8	14	2	248	0	2	4	4
PARTY	0	1	0	2	5	96	0	0	1
PEOPLE	0	1	2	0	2	0	11	0	0
SPECPER	0	0	0	0	0	0	1	10	0
UNION	0	1	2	4	0	1	0	2	61

Table 11: Confusion matrix for the manual resolution of referents of ambiguous pronouns in parliamentary debates (A1: Annotator 1, A2: Annotator 2).

Name	Value
attention_probs_dropout_prob	0.1
hidden_act	gelu
hidden_dropout_prob	0.1
hidden_size	768
layer_norm_eps	1e-12
max_position_embeddings	512
num_attention_heads	12
num_hidden_layers	12
transformers_version	4.6.1
type_vocab_size	2
vocab_size	31102

Table 12: Parameters/settings used in our experiments.