ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 277

Who are They Talking About? Detecting Mentions of Social Groups in Political Texts with Supervised Learning

Hauke Licht Ronja Sczepanksi

February 2024 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

Who are they talking about? Detecting mentions of social groups in political texts with supervised learning

Hauke Licht* Ronja Sczepanski†

February 6, 2024

Abstract

Politicians appeal to social groups to court their electoral support. However, quantifying which groups politicians refer to, claim to represent, or address in their public communication presents researchers with challenges. We propose a novel supervised learning approach for extracting group mentions in political texts. We first collect human annotations to determine the exact text passages that refer to social groups. We then fine-tune a Transformer language model for contextualized supervised classification at the word level. Applied to unlabeled texts, our approach enables researchers to automatically detect and extract word spans that contain group mentions. We illustrate our approach in three applications, generating new empirical insights how British parties use social groups in their rhetoric. Our methodological innovation allows to detect and extract mentions of social groups from various sources of texts, creating new possibilities for empirical research in political science.

Keywords: social groups, political rhetoric, computational text analysis, supervised classification

8578 words (excluding the reference section and appendices)

*University of Cologne, Cologne Center for Comparative Politics, hauke.licht@wiso.uni-koeln.de

†Sciences Po Paris, Center for European Studies and Comparative Research, ronja.sczepanski@sciencespo.fr

1 Introduction

Social groups' struggle to influence political processes and outcomes shapes politics worldwide. Understanding the role of social groups in politics is thus a central theme in many fields of political science research, ranging from political sociology to conflict studies. It is thus not surprising that the extant political science literature offers many hypotheses about how and why politicians relate themselves to social groups or talk about them in their public communication (e.g., Chandra, 2012; Huber, 2021; Kitschelt, 2000; Lieberman and Miller, 2021; Saward, 2006; Stückelberger and Tresch, 2022; Thau, 2019). Yet, quantitatively studying this facet of politics is currently limited by a lack of scalable measurement instruments allowing researchers to quantify group-based political rhetoric.

This paper proposes a novel supervised text classification strategy for extracting social group mentions from large political text corpora. We task human coders with marking all passages in a sample of sentences that mention social groups. This first step results in a set of labeled sentences in which a varying number of words are marked as containing mentions of groups. We then use these annotations to fine-tune a Transformer-based supervised token classifier. The classifier learns to predict whether or not a word in a sentence belongs to a social group mention while accounting for the word's surrounding sentence context. The resulting classifier thus automates our manual word-level annotation procedure and enables reliable detection and inductive discovery of group mentions in unlabeled texts.

We demonstrate the reliability and validity of our method in analyses of the British parties' group-based rhetoric. Our approach proves very reliable in detecting mentions of social groups – even in social group references not contained in the training data or when applied to German party manifestos and British parliamentary speeches. Further, our evidence underscores the validity of our approach. Our document-level indicators of social groups' salience in parties' manifestos we obtained with our approach correlate strongly positively with comparable indicators obtained through fully manual content analysis, and the resulting time series align with descriptive results by Thau (2019).

To illustrate the added value of our method, we apply it in three analyses. First, we demonstrate that parties tend to emphasize social groups more when they discuss (re)distributive policy issues in their manifestos compared to regulatory policy issues. Further, this tendency is pronounced more strongly for the British Labour Party than the Conservative Party in policy areas such as social welfare. Second, we apply our method to reveal differences in the words and phrases that distinguish British parties' social group mentions. This analysis shows that, for example, Labour strongly emphasizes working-class people, the poor, and age-based groups, whereas the Tories emphasize white-collar occupational groups, agricultural workers, immigrants, and criminals. Third, we show that sentences mentioning social groups are more emotional in tone than sentences without such mentions, suggesting that these two rhetorical strategies tend to be linked in parties' campaign communication.

Our innovation thus equips researchers with new flexibility in their analyses of social groups' role in political rhetoric. For example, it will allow them to complement existing studies that focus on how voters respond to group-based political rhetoric (Hersh and Schaffner, 2013; Holman et al., 2015; Robison et al., 2021; Weber and Thornton, 2012) with new studies examining whether and how politicians use these as part of their electoral strategies (e.g., Stückelberger and Tresch, 2022; Thau, 2021). Moreover, because our approach allows locating where social groups are mentioned in a text, researchers can study differences in how politicians talk about specific target groups (e.g., refugees, women, the unemployed, ethnic minorities, etc.). This will facilitate new research into stereotyping, blame attribution, and elites' role in constructing and priming social identities.

2 Social groups in political rhetoric

Social groups are at the heart of political science theory. However, scholars disagree to some extent about how to conceptualize a social group. Some limit their conception of a social group to only include collectives of people that share socio-economic circumstances or socio-demographic characteristics (Dolinsky et al., 2023; Huber, 2021) that provide a source of identification for group members (Miller et al., 1991). Others, like Howe et al. (2022), advocate for a more open conception, arguing from a constructivist perspective (Chandra, 2012; Wolkenstein and Wratil, 2021) that a social group can be any collective of people that share some attribute, including common values and life experiences. For example, attributes like "hard-working" and "moral righteousness" can be central to people's conceptions of their in- and out-groups (Sczepanski, 2023; Zollinger, 2022).

These differences in conceptualizations have important implications. Both conceptions consider collectives with shared economic circumstances and socio-demographic characteristics as social groups, such as "workers," "families," "pensioneers," or "students." However, unlike scholars with a more inclusive definition, researchers adopting the definition based on socio-structural characteristics would not consider mentions such as "the honest people" or "those who work hard" as social group references. One of the key distinctions between the two conceptions thus is what kind of boundaries they focus on (Mierke-Zatwarnicki, 2023). The socio-economic definition focuses on boundary drawing in line with the distribution of material resources and 'objective' demographic characteristics. In contrast, more abstract group references also focus on symbolic, discursively constructed boundaries such as "honest people" (Lamont and Molnár, 2002; Mierke-Zatwarnicki, 2023).

We opt for the broader, more inclusive conceptualization. As we study the reference to social group categories in political speech and text, we cannot apply group members' identification as a criterion. More importantly, even symbolic boundaries can turn into social boundaries and eventually political cleavages if they are politicized (cf. Enyedi, 2005). By capturing all social categories that might turn into meaningful social and political boundaries, we thus take into account politicians' agency to construct groups.

Politicians have many reasons to emphasize social groups by directly referring to them in their public communication (Conover, 1988; Miller et al., 1991). Talking more or less about social groups allows parties and their representatives to show which groups are important to them and which are not (Conover, 1988; Dolinsky et al., 2023; Gadjanova, 2015; Horn et al., 2021; Howe et al., 2022; Nteta and Schaffner, 2013; Stückelberger and Tresch, 2022; Thau, 2019). Mentioning a social group frequently can be a way to signal responsiveness to it and make its members "feel seen" and represented in politics (Pitkin, 1967; Robison et al., 2021; Saward, 2006). Further, emphasizing social groups in their public communication can allow politicians to mobilize groups' sentiments, identities, and grievances (Goodman and Bagg, 2022; Miller et al., 1991; Stückelberger and Tresch, 2022).

But group-based rhetoric is also about shaping groups' opinions, interests, and perceptions (Enyedi, 2005; Wolkenstein, 2021). For example, how elites talk about social groups can affect how positively or negatively these groups are viewed by others – often with consequences for how deservingness perceptions (O'Grady, 2022; Slothuus, 2007). Thus, political parties and their representatives can shape groups' standing in society. Moreover, research has shown that connecting groups to an issue position can alter their opinion on the topic (Huber et al., n.d.). Therefore, which groups politicians appeal to can also affect how citizens perceive their political and social world.

In sum, elites' group-based rhetoric potentially shapes how represented citizens feel, what shared interests they perceive, how they feel towards other groups, and their opinions about political issues. It is thus of central interest to political scientists to understand when, why, and how politicians mention social groups in their public communication.

Yet, studies of political elites' group-based rhetoric are still relatively rare. A lot of research has focused on citizens' perceptions of group appeals and their feelings of being represented as a group (Holman et al., 2015; Jackson, 2011; Kam et al., 2017; Robison et al., 2021; Valenzuela and Michelson, 2016; White, 2007). In contrast, research on the "supply" of group-based rhetoric is currently largely limited to a handful of studies in the party politics literature (e.g., Dolinsky, 2022; Horn et al., 2021; Howe et al., 2022; Huber, 2021; Stückelberger and Tresch, 2022; Thau, 2019, 2021) and research on ethnic politics (e.g., Lieberman and Miller, 2021; Nteta and Schaffner, 2013). One of the main challenges to studying social groups in political texts is to detect them in large amounts of texts and across contexts.

3 Detecting mentions of social groups in political texts

We argue that one of the main reasons comparative research on political actors' use of group-based rhetoric is limited in scope lies in the methodological challenges researchers confront when trying to detect and extract social group mentions in large political text corpora. As outlined next, these challenges are largely due to social group mentions' linguistic characteristics. These characteristics, in turn, limit the reliability and scalability of existing content-analytic measurement approaches. The supervised token classification approach to group mention detecting we introduce overcomes these challenges.

Characteristics of group mentions in political texts

One of the foremost methodological challenges in identifying mentions of social groups in political text and speech is that they are linguistically extremely diverse. First, the number of social groups that can be referred to in a given political context is typically large. The list is already long if one considers only groups that are defined based on socio-demographic characteristics like age or generation, gender, race, or ethnicity (cf. Chandra, 2012). And if one considers that objective membership in different group categories is often nested and intersectional, the list grows further. For example, a mention of "people living and working in rural areas" refers to members of the rural population who are workers.

Second, political actors do not only refer to groups using socio-demographic markers but also discursively construct groups by emphasizing people's shared values, norms, circumstances, and commonalities in other attributes. Research in political psychology (Huddy, 2001) and the study of representation (Wolkenstein, 2021; Wolkenstein and Wratil, 2021) suggests that politicians invoke groups discursively to produce in- and outgroup affect (cf.

Table 1. Examples of group mentions in sentences drawn from British mainstream party manifestos. Highlighted text spans identify groups mentioned in each sentence.

`We seek to bring about a fundamental change in the balance of power and wealth in favour of working people and their families.`
`Eight years of meanness towards the needy in our country and towards the wretched of the world.`
`The welfare of the old, the sick, the handicapped and the deprived has also suffered under Labour.`
`Labour recognises the special needs of people who live and work in rural areas.`

Hobolt et al., 2021; Zuber et al., 2023). For example, phrases like "the needy in our country" and "the wretched of the world" (see Table 1), "those with the broadest shoulders," or "those who work hard and do the right thing" refer to no clearly circumscribed socio-demographic groups, but still likely appeal strongly to people with corresponding self-conceptions and identities (Bornschier et al., 2021).

A third reason social group mentions in political texts are linguistically extremely varied is that for any given social group, there are various lexically different ways to refer to it. For one, there are many indirect ways to refer to a group. For example, the phrases "the unemployed" and "those out of work" refer to the same social group. For another, many references to groups use descriptive language, such as "the first generation to know we are destroying the environment, and the last generation with a chance to do something about it before it is too late."

The linguistic diversity of social group mentions in political rhetoric has two important methodological implications. First, the phrases used to mention, refer to, or address social groups in political text often span multiple words. Second, any sentence can mention no, one, or several social groups (see Table 1 for examples). Accordingly, reliable detection and extraction of social group mentions requires identifying the words used to refer to or invoke social groups in a text while not knowing *a priori* how many unique mentions it contains, where the mentions are located in the text, and how many words a given mention spans.

Established approaches and their limitations

To cope with these challenges, researchers studying groups-based rhetoric based on political text currently have two options: manual content analysis or automated dictionary measurement. These two approaches are well-established in the applications to sentence- and document-level classification (cf. Barberá et al., 2021; Quinn et al., 2010). However, both approaches have clear limitations when applied to extract group mentions from large text corpora.

Manual content analysis identifies group mentions in political texts by tasking coders to locate and extract the relevant text segments referring to groups (e.g., Huber, 2021; Stückelberger and Tresch, 2022; Thau, 2019, 2021) or by indicating this information at the sentence level (Hopkins et al., 2022; Horn et al., 2021). As in other applications (cf. Grimmer and Stewart, 2013; Quinn et al., 2010), this approach can be considered the most valid one. Human coders can read and interpret texts, allowing them to spot simple group mentions as well as more complex ones, like the abstract or descriptive multi-word examples included in Table 1 above.

However, when deciding to task their coders to only indicate whether or not a sentence contains one or more references to a group (category), researchers gain efficiency but miss the opportunity to record group mentions' exact wording. Discarding this information during the manual coding process limits researchers' ability to gain more detailed knowledge, for example, about how exactly politicians appeal to groups. Moreover, it prevents them from inductively organizing the extracted group mentions into broader group categories (e.g., references to 'economic' vs. 'non-economic' groups, as in Thau, 2019), as has been key in, for example, the studies by Thau (2019), Huber (2021), and Stückelberger and Tresch (2022).

More importantly, manual content analysis is relatively costly (but see Benoit et al., 2016), as researchers need to hire annotators, and it takes considerable time to collect annotations for all sentences in large corpora. Consequently, studies that have applied manual content analysis to study group-related rhetoric use either text corpora of limited size, focus on a small set of political parties, and/or limited periods.

The dictionary approach is more resource-efficient compared to manual content analysis as dictionary-based measurement enables detecting mentions of pre-defined groups automatically by searching for matches to a list of group keywords (cf. Dolinsky et al., 2023). The only required human input to dictionary-based measurement is to define a list of keywords that reflect the potential ways the social group(s) of interest are mentioned in a corpus.

However, considering how linguistically varied social group references are in political texts, we should expect that compiling a comprehensive list of relevant keywords will be very challenging in many applications. Even for experts, it can be difficult to anticipate all potential ways a target group is mentioned in a corpus before seeing the data, especially since group mentions usually span multiple words, are often indirect, and potentially discursively invoke groups in abstract ways. For example, a group-specific dictionary may detect "the unemployed" but fail to recognize semantically similar phrases like "those out of work."[1]

To summarize, manual content analysis allows valid measurement of social group mentions in political texts but is resource-intensive and when adopting a sentence-level classification approach, it means discarding empirically interesting variation. Dictionary analysis, in contrast, is very resource-efficient but limits reliable detection, especially for groups without clear-cut membership criteria and groups that can be referred to in many lexically different ways.

[1]Searching for relevant keywords iteratively (cf. Muddiman et al., 2019) risk over-fitting to the subset of the corpus the researcher has reviewed to build their dictionary and thus limit generalization. Including only indicator words (here, e.g., "those" and "work") would lead to many false-positive classifications. Checking for co-occurrences of such words in documents (e.g., "those" + "work", "those" + . . .) could partially remedy this concern. However, the number of keywords that require inclusion increases rapidly with the length of relevant expressions, while increasing the number of keywords in a dictionary often reduces precision due to polysemy.

Figure 1. From sentence annotation to extracted mention. Highlighted spans are converted into token-level labels. Labels ‘`B`’ and ‘`I`’ indicate tokens that are at the “beginning” or “inside,” the ‘`O`’ those outside of a group mention. The token classifier predicts label probabilities, which indicate a token’s most likely label. Predicted mentions can be determined from token-level predicted labels.

Input data (text with group mention annotations)

`"We represent those who work hard and do the right thing."`

$\Downarrow$

Model inputs (tokenized text and corresponding token-level labels)

tokens	We	represent	those	who	work	hard	and	do	the	right	thing	.
labels	O	O	B	I	I	I	I	I	I	I	I	O

$\Downarrow$

Model outputs (token-level predicted probabilities and labels)

tokens	We	represent	those	who	work	hard	and	do	the	right	thing	.
probabilities												
O	0.58	0.49	0.30	0.30	0.30	0.31	0.32	0.31	0.31	0.31	0.32	0.63
I	0.22	0.24	0.28	0.45	0.47	0.47	0.47	0.46	0.47	0.47	0.48	0.20
B	0.20	0.28	0.43	0.25	0.23	0.22	0.21	0.22	0.22	0.22	0.20	0.17
labels	O	O	B	I	I	I	I	I	I	I	I	O

$\Downarrow$

Post-processed outputs (extracted group mentions)

span `"those who work hard and do the right thing"`

A supervised token classification approach

We propose a method that allows researchers to automatically identify and extract mentions of groups in political texts with a limited amount of manual labeling effort. Our method applies supervised learning to detect and extract mentions of social groups in political texts. It strikes a favorable balance between the objectives of reliable and valid detection on the one hand and scalability on the other.

The first step of our supervised learning approach is to task human coders to highlight all mentions of social groups in a set of sentences sampled from a target corpus. This step mirrors the procedures adopted in existing manual content analysis studies. However, what distinguishes our approach is that we preserve the verbatim mentions of groups where and how they occur in texts.[2] The first row in Figure 1 illustrates what the annotations we collect look like. By tasking coders with highlighting all group mentions in a sentence, we can determine the characters that belong to individual group mentions.This means that in each labeled sentence, no, one, or several spans of characters might be marked as mentioning a group (see Table 1 for examples).

In the second step, we use this information as data for supervised learning. Specifically, we train a supervised classifier for token classification. Token classification means to assign each word in a sentence a single label from some pre-defined categories. Enabling this requires converting the annotations into word-level labels. This is illustrated in the second panel of Figure 1. From the annotations we have collected in the first step, we know for each group mention in a sentence at which character it starts and ends. Tokenizing the sentence into words, we can determine for each word in the sentence whether or not it belongs to a mention of a group. Further, for words that belong to such a mention, we can determine whether the word is at the beginning of the mention or inside of it. As shown in the second row of Table 1, words that do not belong to a mention are labeled '`O`' to indicate that they are *outside* of a social group mention. In contrast, words at the beginning or inside of a mention are labeled '`B`' respectively '`I`' (cf. Ramshaw and Marcus, 1995).

With word-level labels at hand, the supervised token classification task is to predict each word's label in a sentence. Provided with multiple labeled sentences in this format, we fine-tune a Transformer-based neural network for this task. Relying on a pre-trained Transformer-based model like BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019) or DeBERTa (He et al., 2021) allows accounting for words' sentence context when learning to predict their labels. This is impossible with standard bag-of-words methods.

The result of this second step is a fine-tuned token classification model that can be applied to detect and extract mentions of social groups in political texts. As shown in the third panel in Figure 1, the label class that receives the highest predicted probability for a

[2] For example, this contrasts with Thau (2019), who has "cleaned" and "harmonizing" (i.e., post-processed) the real-occurring mentions.

word is treated as its predicted label. And as shown in the last panel of Table 1, this allows extracting the words belonging to the (predicted) group mention.

In the third step, the fine-tuned supervised token classifier can be applied to unlabeled texts to identify and extract mentions of social groups that have not been in the training data. This enables automated labeling and extraction of group mentions in large text corpora.

Our proposed method thus contrasts in three important ways with established approaches to quantifying group-based rhetoric in political texts. First, it contrasts with dictionary-based measurement in that we presume that recognizing concrete group mentions in a text is more reliable than selecting indicative words or phrases *a priori*. Second, in contrast to the manual content analysis approach, we leverage the benefits of automation through supervised learning. This saves researchers the time and costs associated with classical manual content analysis (cf. Barberá et al., 2021; Grimmer and Stewart, 2013). Third, in contrast to sentence-level classification approaches, we annotate, model, and predict the text passages that represent group mentions at the word level. Consequently, our approach preserves the lexical diversity and linguistic variability of group mentions as they occur in political texts, which will enable more detailed analyses of group-centered political rhetoric.

4 Evaluating our approach in British party manifestos

To validate our method, we focus on detecting and extracting social groups mentioned in British parties' election manifestos in this paper and report additional analyses of social group mentions in parliamentary questions in the UK *House of Commons* as well as in German parties manifestos in the appendix (see Appendices A and E.3). Our case selection is motivated by substantive as well as methodological considerations. From a substantive perspective, we are interested in comparing parties' social group mentions across elections and parties. From a methodological point of view, studying cases that in parts have already been studied, like the UK Labour and Conservatives Party manifestos between 1964 and

2015 (Thau, 2019), enables us to evaluate whether our results align with ones derived from manual content analysis.

Data and methods

Our primary dataset records 28 election manifestos of the two largest British parties – the Labour Party and the Conservative Party – from the elections of 1964 to 2019. We complement this dataset with the manifestos of the Democratic Unionist Party (DUP), Green Party of England and Wales (Greens), Liberal Democrats (LibDem), Scottish National Party (SNP), and United Kingdom Independence Party (UKIP) for the elections 2015, 2017, and 2019. Our parliamentary question time and German party manifest data sets are described in Appendix A. We have split manifestos' raw texts into sentences (see Table A1) and drawn a sample of 8,596 sentences from this corpus for annotation, stratifying by party and (election) year and, where possible, by manifesto chapter (see Table B4).[3]

To collect annotations of social group mentions in these documents, we have designed a custom coding scheme. The focal category of our coding scheme is the "social group" category. In our application, we define "social group" as a collective of people with one or multiple common characteristics. As discussed in section 2, we deliberately adopt a broad conceptualization. In addition, we include four other categories in our coding scheme ("political group", "political institution", "organization etc.", and "implicit social group reference;" see Table B6 in the Appendix) and an "unsure" category. We included these additional categories for three reasons. First, when developing the coding scheme, we found that additional categories helped our annotators to recognize the conceptual boundaries of the "social group" category. Second, collecting annotations for these categories allows us to demonstrate that our method is reliable also in detecting other types of groups. Third, we wanted to make our data as reusable for secondary research as possible.

[3]This sampling strategy ensures that data from all election years and parties are represented equally in our training data. Stratifying by manifesto chapter moreover enhances the topical coverage of our labeled dataset.

We have collected annotations from two trained research assistants using the `doccano` online annotation tool (Nakayama et al., 2018). As shown in Table B5 in the Appendix, we have collected annotations from both coders for more than 30% of sentences because it is a well-known limitation of content-analytic annotation procedures like ours that individual coders can make mistakes or some text passages might be ambiguous (cf. Krippendorff, 2004). As shown in Table B8, the inter-coder agreement is very high in our sample of doubly annotated sentences. The median (mean) sentence-level agreement in sentences with at least one social group annotation by either coder is 95.7% (90.8%) and 95.2% (91.5%) in sentences without any social group annotation but at least one other group annotation. This indicates that our coding instrument and procedure indeed elicit highly reliable annotations. Moreover, analyzing the sentences with disagreements, we find that in a sizeable number of sentences (24-45%), our coders' disagreements stem from mismatches in the exact beginning, end, or beginning and end of individual group mentions (see Table B9).

Because we have collected annotations from two coders for some sentences, we need to aggregate these annotations into a single set of word-level labels per sentence. As described in Appendix B.1, we follow the rich computer science literature on annotation aggregation (cf. Chatterjee et al., 2019) and fit a Bayesian sequence combination model (Simpson and Gurevych, 2019). This results in word-level labels for all 8,576 human-annotated sentences in our annotated British manifesto sentences.

We have used the resulting labeled sentences to fine-tune RoBERTa models (Liu et al., 2019) for token classification.[4] To prepare the labeled data, we have first dropped all "unsure" annotations so that the corresponding words are treated as if they are not part of any type of group mention. We have then converted sentences' word-level labels into the IOB2 (inside–outside–beginning) label scheme (Ramshaw and Marcus, 1995). This means tokens at the beginning of a mention receive a special label. For example, we distinguish between tokens at the beginning of social group mentions (B-SG) and tokens inside them (I-SG). Together

[4]We have run an experiment to compare the performance when using other pre-trained models (BERT, DistillBERT, and DeBERTa v3) but found no improvements.

with the "outside" (O) label reserved for tokens outside of a mention, this results in eleven label classes.

Results

Reliability: Out-of-sample classification performance

In Table 2, we report the results of 5-times-repeated 5-fold cross-validations of token classifiers fine-tuned on labeled sentences in our UK party manifesto corpus.[5] Cross-validation allows us to summarize the results of 25 different classifiers, and we can thus present robust estimates of classifiers' out-of-sample performance.[6]

Focusing on classifiers' reliability in detecting social group mentions,[7] we first turn to their average mention-level performance. We compute mention-level recall, precision, and the F1 scores estimates by comparing predicted to "true" word-level labels within observed and predicted group mentions and averaging these estimates across social group mentions in the test set.[8] Looking at classifiers' performance at the mention level, they correctly classify on average 93% of words that belong to social group mentions in the human-labeled data (recall). Conversely, our classifiers are correct 94% of the time when they predict that a word belongs to a social group mention (precision). This amounts to an average mention-level F1 score of 96%.

This high level of reliability in detecting social group mentions in held-out texts translates into extremely high sentence-level performance. To compute sentence-level performance from

[5]We have iterated over five random seeds to control the initial train/test split and then iterated in a 5-fold splitting over the training data to train five different classifiers per random seed on different train/development splits.

[6]Note that we have grouped by manifesto when splitting the data to prevent data leakage and increase the ecological validity of our analysis. This means all labeled sentences from a manifesto are either in the training, validation, or test set. Depending on the random seed, this approach resulted in training sets with 6,108 to 6,245 labeled sentences, validation sets with 809 to 896 labeled sentences, and test sets with 1,480 to 1,574 labeled sentences.

[7] C12 reports the results for all group categories.

[8]Say we have a span like "the British people" with true labels [`B-SG`, `I-SG`, `I-SG`]. If our classifier correctly predicts the labels of all words in this span, recall is 1.0. If, however, it misses the first word (e.g., it predicts [`O`, `B-SG`, `I-SG`], recall is 0.666 (or 2 out of 3 correct, ignoring the distinction between inside and beginning category types).

Table 2. Summary of test set performances of our group mention detection classifier trained and evaluated on our corpus of labeled UK manifesto sentences. Values (in brackets) report the average (90% quantile range) of performances of 25 different classifiers trained in a 5-times repeated 5-fold cross-validation scheme. Rows distinguish between different evaluation schemes (i.e., different ways to compute the F1 score). *Note:* `seqeval` is the strict metric proposed by Ramshaw and Marcus (1995).

	F1	Precision	Recall
Mention level	0.96 [0.94, 0.99]	0.94 [0.90, 0.97]	0.93 [0.90, 0.98]
Sentence level	0.97 [0.94, 0.99]	0.97 [0.94, 0.99]	0.97 [0.94, 0.99]
`seqeval` metric	0.87 [0.80, 0.93]	0.86 [0.79, 0.93]	0.88 [0.81, 0.94]

word-level predictions, we determine for each group category in our coding scheme whether there is at least one annotation in the "true" and predicted labels, respectively, and compare them within sentences. We then count a sentence as correctly classified if, for the given group type, at least one word was labeled correctly. According to this standard, in sentences that contain at least one social group mention, our classifiers correctly classify on average 97% of sentences (recall). In expectation, this amounts to only three misclassifications per 100 sentences that contain one or more social group mentions.

Table 2 also reports the so-called `seqeval` metric, which considers mention-level predictions only as correct if the classifier predicts the correct label for *each* word in a given human-labeled mention. Instances where the classifier's prediction begins too late or early, ends too early or late, etc., are considered classification errors. Even according to this rather strict standard, our classifiers correctly predict 88% of social group mentions (recall), 86% of the social group mentions they predict are correct (precision), and this amounts to an average F1 score of 0.87. We note, however, that based on our review of our coders' annotations, minor disagreements on the exact beginning or end of group mentions are often inconsequential for capturing the essence of true group mentions. The strict standard the `seqeval` metric applies thus arguably results in overly conservative classification reliability estimates.

The out-of-sample classification performances reported in Table 2 indicate that our supervised token classification approach to social group mention detection yields highly reliable measurements. In the Appendix, we report additional evidence that supports this conclusion. First, the classifiers evaluated in Table 2 achieve similar levels of reliability in the other group types included in our coding scheme (see Table C12). Second, assessing the effect of the number of training samples on out-of-sample classification performance, we find that similar levels of reliability as those reported in Table 2 can be achieved when training on only 3,000-4,000 labeled sentences (see Appendix E.1). Third, we present evidence that our classifiers generalize well, as they can reliably detect social group mentions not contained in their training data (see Appendix E.2). Fourth, we show in Appendix E.3 that our classifiers trained on British party manifestos can be transferred reliably to a different domain (parliamentary speech) and another language (German party manifestos) with very little additional labeled data (cf. Ho and Chan, 2023; Licht, 2023).

Validity: Comparison to measurements obtained by Thau (2019)

We next demonstrate that the measurements generated with our approach also converge with the measurements obtained by Thau (2019). Thau (2019) has tasked trained coders with manually coding group-based appeals made in UK Labour and Conservative party manifestos (1964–2015). Part of this task is identifying the explicit mentions of targeted social groups.

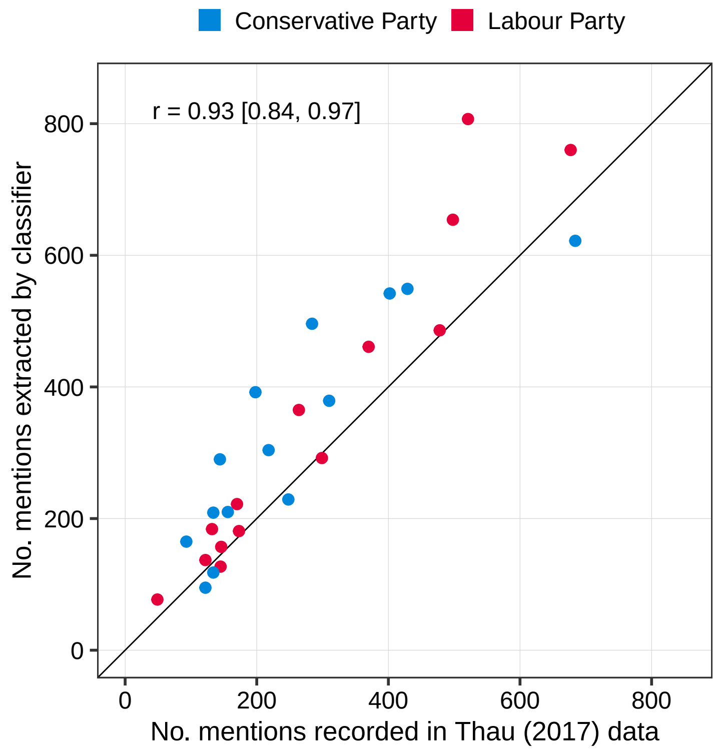
We use Thau's data to validate our approach in two ways. First, we assess whether the social group mentions Thau's coders have identified are also detected by our supervised token classification approach. To answer this questions, To we have matched the group mentions extracted by Thau's coders to the manifesto sentences from which they were retrieved,[9] applied our supervised token classifier to them,[10] and computed average mention-level recall values for each group category in Thau's coding scheme.[11] As shown in Figure D1, our

[9] We describe this procedure in the Appendix.

[10] We have trained this token classifier on 80% of labeled sentences sampled from all UK parties' manifestos.

[11] We focus on recall because Thau has coded group-based appeals. A group-based appeal implies a group mention but not vice versa. Hence, our classifier might detect mentions outside of group-based appeals.

Figure 2. Cross validation of classifier's predictions against data collected by Thau (2019). Figure compares the numbers of social group mentions identified in a manifesto by Thau (2019, see x-axis) and our classifier (y-axis) in Labour and Conservative party manifestos (1964-2015). Colors indicate parties. Correlation coefficient (with 95% confidence interval) shown in top-left of plot panel.

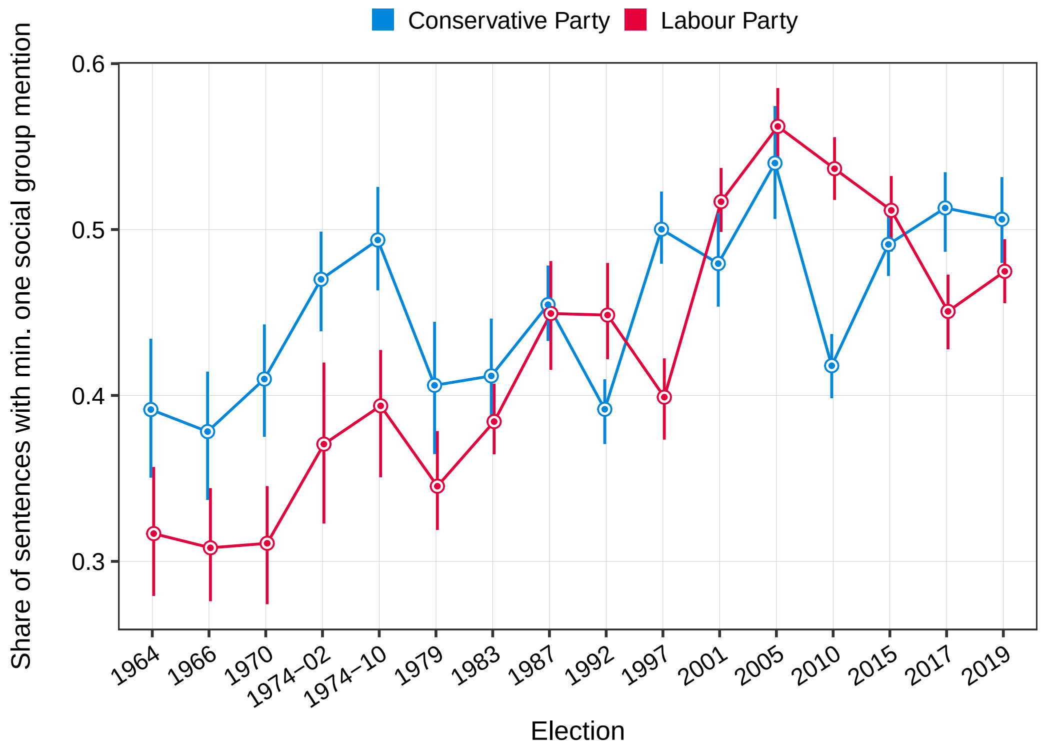


classifier performs overall consistently across his group categories, achieving average mention-level recall values above 0.90 in most categories. As discussed in greater detail in Appendix D, the three exceptions to this pattern are explained by how our coding instructions diverge from Thau's.

Second, we use Thau's data to compare our automated to his manual approach at the document level. Specifically, we count the number of social group mentions in each party manifesto according to his records and our classifier's predictions and compare how they correspond. Figure 2 shows a high positive correlation between our and Thau's estimates. Moreover, our counts are systematically higher, which is expected since Thau has coded group-based appeals, and a group-based appeal implies a group mention but not *vice versa*.

As an additional validation check, we report in Figure 3 how the share of sentences in

Figure 3. Share of sentences in Labour and Conservative Party manifestos that contain at least one social group mention by election. *Note:* To quantify the uncertainty in these estimates, we have bootstrapped sentence-level indicators.



the Conservative and Labour parties' election manifestos that mention at least one social group has changed since the 1964s according to our estimates. It shows that both parties have increased the group-centeredness of their campaign communication until around 2005.[12] Between 1966 and early 1974, the Conservative Party under Edward Heath emphasized social groups more strongly than Labour, and both parties' social group emphases have developed very similarly since 1975. However, our estimates for the election years 2005 and after indicate that the Tories have reduced their social group salience by about ten percentage points. This change of strategy coincides with the beginning of David Cameron's leadership.

[12]Having computed these estimates as the share of sentences, we can rule out that this finding is an artifact of manifestos increasing length over time. Moreover, when sampling sentences from human annotations, we stratified them by party and election. It is thus very unlikely that the over-time increase shown in Figure 3 is an artifact of a temporal bias in our classifier's predictions.

5 Applications

To illustrate the added value of our approach, this section presents three substantively motivated analyses of the measurements we have generated for British parties' election manifestos. These analyses show that our automated social group mention detection and extraction method allows testing theoretical claims and generating novel empirical insights. We first investigate in which policy topics parties talk the most about social groups. Second, we study substantive differences in British parties' social group focus, focusing on what distinguishes the groups they mention. Lastly, we show that sentences that contain mentions of social groups are more likely to include emotional language than sentences without group mentions.

In which policy areas are social groups made most salient?

We first examine how political parties connect policy issues to references of social groups (cf. Huber et al., n.d.; Robison et al., 2021; Thau, 2023). Since group mentions can reflect parties' attempts at addressing groups' interests and shaping their opinions, we expect more mentions in policy areas marked by distributive conflict. Accordingly, we hypothesize that discussions about (re)distributive policies, such as social welfare, will include more social group mentions than discussions about regulatory issues like the economy (Majone, 1997).

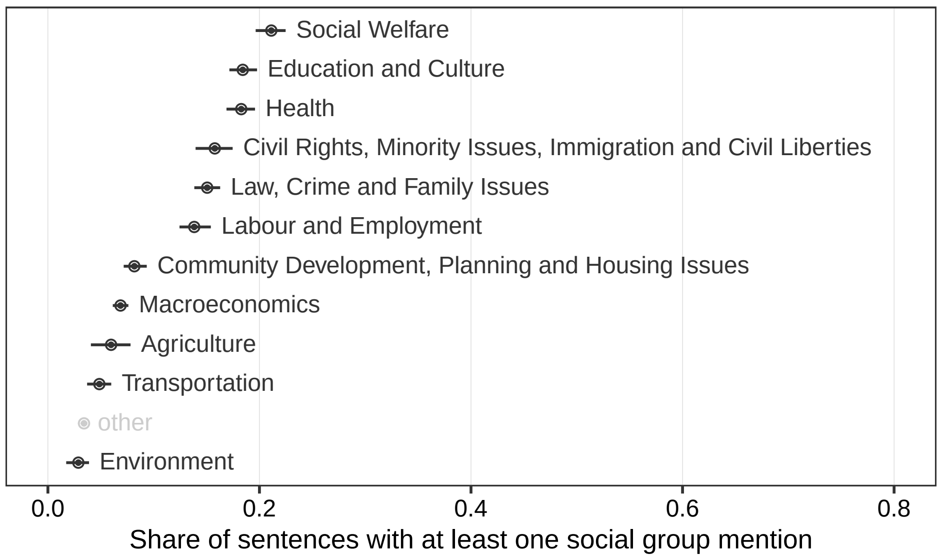
We test this expectation using data from the UK *Comparative Agendas Project* (CAP; Jennings et al., 2011). Specifically, we classify manifesto sentences according to the CAP policy topic they discuss[13] and then estimate the prevalence of social group mentions in sentences by CAP category.

Figure 4 reports the share of sentences that mention at least one social group by CAP policy topic.[14] The salience of social group mentions in different policy topics aligns with

[13] We fine-tune a policy topic classifier with human-coded quasi-sentences from UK Labour and Conservative party manifestos (1983-2015) in the CAP data.

[14] Note that we have collapsed the topics 8 ("Energy"), 15 ("Banking, Finance and Domestic Commerce"), 16 ("Defence"), 17 ("Space, Science, Technology and Communications"), 18 ("Foreign Trade"), 19 ("Interna-

Figure 4. Salience of social group mentions by *Comparative Agendas Project* (CAP) policy topic. Computations based on social group mentions in Labour and Conservative party manifesto sentences (1983-2015). *Note:* Sentences CAP-coded using multiclass classifier trained on human-labeled manifestos of same cases (Jennings et al., 2011) Infrequent CAP policy topics grouped into the "other" category. Topic "Immigration" recoded to topic "Civil Rights, Minority Issues, Immigration and Civil Liberties".

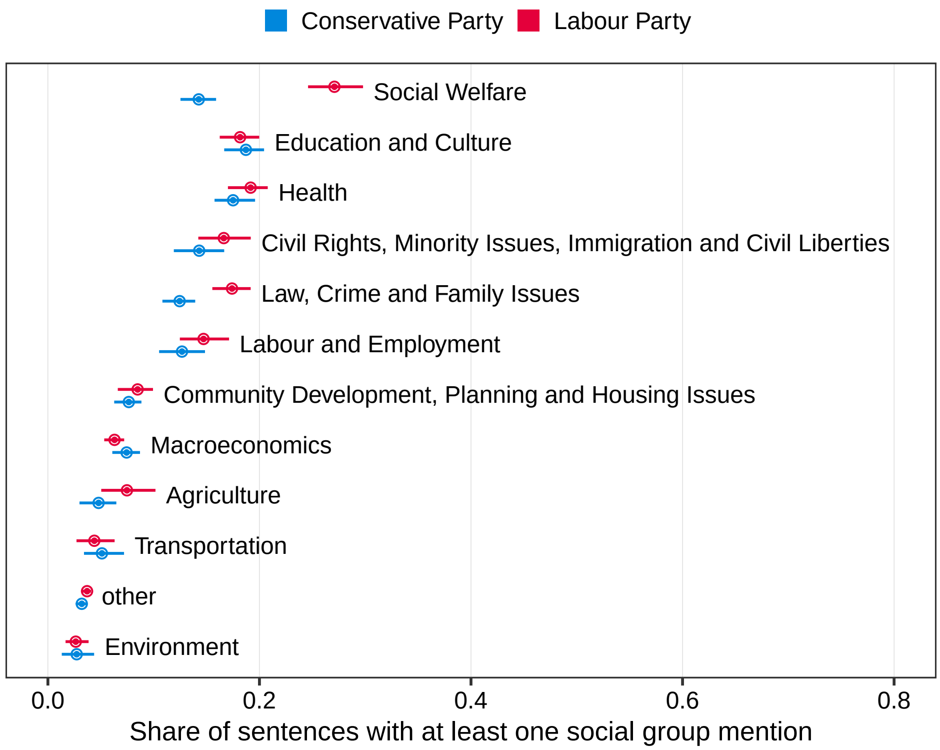


our expectations. Distributive and redistributive policy areas (e.g., social welfare, education, civil rights) are more likely to include social group references than sentences about regulatory matters (e.g., transportation, environment).

Further, as shown in Figure 5, we observe differences between parties in how much they emphasize social groups when addressing these policy issues. Labour mentions social groups more in their manifestos than the Conservatives when talking about the topics of "Social welfare" and "Law, Crime, and Family issues." A reverse pattern emerges tentatively in their discussion of macro-economics topics. This suggests that parties emphasize social groups more in areas considered their core competencies, indicating an association between emphasis on social groups and issue ownership (Petrocik, 1996).

tional Affairs and Foreign Aid"), 20 ("Government Operations"), and 21 ("Public Lands, Water Management, Colonial and Territorial Issues"), into one "other" category because they were extremely sparsely populated. Moreover, we have assigned sentences originally coded into to the "Immigration" topic to the "Civil Rights, Minority Issues, Immigration and Civil Liberties" topic because they were very imbalanced across parties.

Figure 5. Salience of social group mentions in Labour and Conservative party manifestos (1983-2015) by *Comparative Agendas Project* (CAP) policy topic. *Note:* Sentences CAP-coded using multiclass classifier trained on human-labeled manifestos of same cases (Jennings et al., 2011) Infrequent CAP policy topics grouped into the "other" category. Topic "Immigration" recoded to topic "Civil Rights, Minority Issues, Immigration and Civil Liberties".

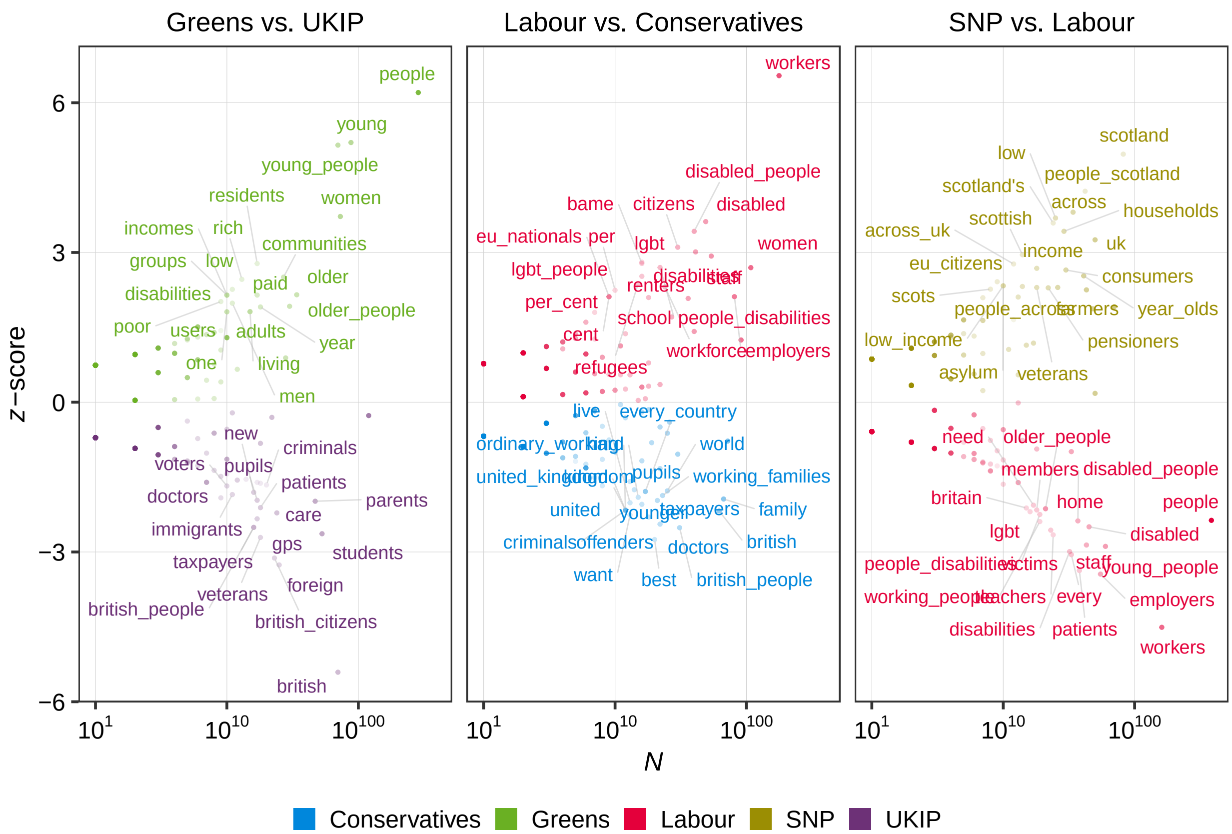


What distinguishes parties' group focus?

Next, we analyze how British political parties distinguish themselves through their references to social groups. Previous studies emphasize that it is not only important *whether* groups are mentioned but also *which* groups (Huber, 2021; Thau, 2021) and *how* they are referred to (Graf et al., 2023).

To investigate this, we employ the "fightin' words" method by Monroe et al. (2008) to the social group mentions identified and extracted by our classifier, focusing on manifestos from 2015 to 2019 to allow the inclusion of smaller British parties in our analysis. The "fightin' words" algorithm is a bag-of-words method for quantifying differences in word choices between speakers, parties, or any other binary indicator. We use this method to

Figure 6. Comparisons of different pairs of parties in terms of the words and phrases that distinguish the social groups the mention in their manifestos for the elections 2015, 2017, and 2019. *Note:* z-scores indicate words "distinctiveness" and have been obtained by applying the "fightin' words" method proposed by Monroe et al. (2008) to the social group mentions retrieved by our classifier.



compare parties' social group mentions between pairs of parties. Specifically, we apply it to the predicted group mentions extracted from parties' manifestos after removing common stop words, retaining uni- and bi-grams, and adding skip-grams.

Figure 6 summarizes our findings. The x-axis shows term frequency. The y-axis displays z-scores that quantify how distinctive the words a party uses to refer to social groups when comparing pairs of parties. Higher z-scores indicate more distinctive words.

Analyzing Conservative and Labour manifestos, Labour emphasizes workers' and disadvantaged groups like people [with] disabilities,' refugees,' women, and "BAME" (Black, Asian, and minority ethnic) and LGBT communities. In contrast, the Conservative Party focuses on ‘ordinary working [people]’, ‘working families,’ ‘British people,’ and the middle class (e.g., ”doctors,” ”entrepreneurs,” and ”professionals”).

Examining Greens and UKIP along the GAL-TAN dimension, Greens refer distinctively to age- and gender-based groups and disadvantaged communities, while UKIP, like the Conservatives, focuses on ‘the nation’ and ‘British people’, also mentioning immigrants and criminals.

Comparing Labour and the SNP, the center-periphery issue of Scottish independence is evident, with the SNP mentioning ‘[the] people [of] scotland’, ‘scottish’, ‘scotland’s’ people, citizens, etc., as well as ‘scots’.

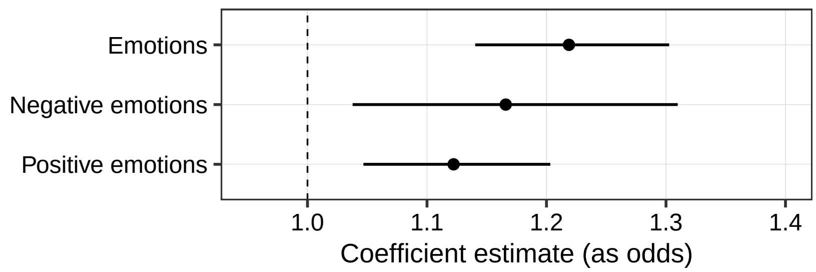
These findings underscore the practical value of our method for researchers. By extracting the exact words with which parties refer to social groups, our method facilitates inductive discovery and analysis of party rhetoric based on a limited set of human-annotated sentences.

Is group-based rhetoric linked to emotional appeals?

Like directly mentioning social groups, emotional language is a powerful rhetorical strategy to appeal to voters (Crabtree et al., 2019; Gennaro and Ash, 2022; Osnabrügge et al., 2021). However, we do not know whether parties combine these two strategies in their campaign communication or use them separately.

We investigate the link between group-based rhetoric and emotional appeals through logistic regression analysis. We use our sentence-level corpus of automatically labeled Labour and Conservative party manifestos from 1964 to 2019. Our dependent variable measures whether a sentence includes emotional language based on the *Linguistic Inquiry and Word Count* (LIWC) dictionary (Pennebaker et al., 2015). Specifically, we classified a sentence as containing emotional language (coded 1) if at least one word matched the list of positive and negative emotion words in the LIWC. If a sentence contains no emotional words, we coded it as zero (0). Further, we have created two additional indicators using only positive and negative emotion words, respectively. These alternative outcomes allow us to assess whether

Figure 7. Coefficients estimates from regressions analyzing whether sentences that contain group mentions are more likely to contain emotion words. The x-axis reports our estimates of the odds that a sentence contains emotional language when it contains at least one social group mention compared to when it contains no social group mention. Points (line ranges) report the coefficients point estimates (95% confidence intervals) of logistic regression models. The y-axis values differentiate between different emotion dictionary categories.



positive, negative, or both emotions contribute to the overall association.

Our main explanatory variable measures whether a sentence mentions one or more social groups and we classify all sentences that contain at least one (predicted) social group mention as 1s and all others as 0s. To account for potential confounders, we control for parties' positions on the economy and cultural topics using Manifesto Project Data indicators (Lehmann et al., 2022), whether a party was the prime minister's party in the year leading up to the election for which the manifesto was written, and the number of words in a sentence. We use these indicators to fit logistic regression models with the binary emotion indicator as the outcome. All our models include election fixed effects, and we cluster standard errors at the level of parties and elections.

Figure 7 presents the coefficient estimates of our regression models for our binary, sentence-level social group mention indicator as odds.[15] The odds measure how much more likely a sentence is to contain emotional words when it contains at least one social group mention compared to when it contains no social group mention. Figure 7 shows that sentences that contain at least one social group mention are about 1.2-1.4 times more likely to contain emotional words. This association exists with positive and negative emotional language use, as

[15] All estimates are reported in Table F15 in the Appendix.

we find positive and statistically significant associations when measuring emotional language use only with positive or negative emotion words in the LIWC dictionary.[16]

This analysis underscores that applying our method for automatically detecting social group mentions in political texts enables new empirical insights into the relation between group-based rhetoric and emotional appeals in parties' campaign communication.

6 Conclusion and discussion

While the extant political science literature offers many hypotheses on how and why politicians relate themselves to social groups in their public communication, studying this facet of politics quantitatively is challenging with existing text-as-data methods. We have proposed a supervised token classification method that enables researchers to automatically identify and extract group mentions in large text corpora based on a small sample of human-annotated documents. Human coders first highlight all text passages that mention social groups in a set of documents sampled from the target corpus. These labeled documents then serve as data to train a supervised token classifier that learns to predict labels at the word level while accounting for words' sentence context. Finally, the resulting classifier allows detecting and extracting group mentions in the entire target corpus.

We have illustrated this method in a study of British parties' group-based rhetoric. Trained on less than 7,000 labeled sentences, our token classifiers prove highly reliable in detecting social group mentions – independent of whether they are evaluated at the sentence or group mention level. Further, our approach yields valid measurements. Document-level indicators of social groups' salience in party manifestos resulting from our supervised token classification approach correlate very strongly with those obtained through fully manual content analysis.

We demonstrated the innovative potential of our method with three different applications.

[16]Additional analyses reported in Table F17 and F18 in the Appendix show that this finding holds when we include minor parties' manifestos in our analysis or focus only on manifestos from the elections of 2015 onwards.

Using our approach to all UK party manifestos in our corpus, we have documented that British parties mention social groups to different extents when discussing different policy topics. Further, our analysis of which words distinguish parties' social group mentions the most uncovered patterns familiar to students of party competition and cleavage formation. Lastly, we have applied our method to study the link between parties' mention of social groups and their use of emotional language, uncovering a strong positive association between these two rhetorical strategies.

Given these results and our encouraging findings of the resource effectiveness, generalization, and transferability of our approach (see Appendix E), we believe that method opens up exciting new avenues for further research. For example, our proposed method could enable analyses of political elites' framing and stereotyping of groups, how they relate different groups to each other, how parties' attempts to create new or maintain existing voter linkages manifest in their communication, and how parties' group-based strategies respond to long-term socio-economic transformations.

We recommend two directions for further methodological research to enable these and other applications. First, future research should focus on developing and testing methods for inductively grouping extracted mentions into conceptually coherent categories (cf. Thau, 2019, p. 70) like those applied in existing manual content analysis (e.g., working-class people, Stückelberger and Tresch, 2022). While our method predicts which parts of a sentence are group mentions, it does not categorize them into types of groups.

Second, we see great potential in our method for closing the gap between the concept of a group *mention* and that of a group *appeal*. To close this gap, researchers will need to measure how politicians relate themselves to the social groups they mention. We believe that existing natural language processing methods, like aspect-based sentiment analysis, would allow learning from labeled data whether a group mentioned in a text is connoted positively or negatively.

Acknowledgments

This project has received funding through the *Center for Comparative and International Studies* of the ETH Zurich and the University of Zurich and the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1 – 390838866. Early versions of this manuscript have been presented at PolMeth Europe 2022, EPSA 2022, and the ECPR Joint Sessions Workshop on Social Groups and Electoral Politics in 2023.

References

Barberá, P., A. E. Boydstun, S. Linn, R. McMahon, and J. Nagler (2021). "Automated Text Classification of News Articles: A Practical Guide". In: *Political Analysis* 29.1, pp. 19–42.

Benoit, K., D. Conway, B. E. Lauderdale, M. Laver, and S. Mikhaylov (2016). "Crowd-sourced Text Analysis: Reproducible and Agile Production of Political Data". In: *American Political Science Review* 110.2, pp. 278–295.

Bornschier, S., S. Häusermann, D. Zollinger, and C. Colombo (2021). "How "Us" and "Them" Relates to Voting Behavior—Social Structure, Social Identities, and Electoral Choice". In: *Comparative Political Studies* 54.12, pp. 2087–2122.

Chandra, K., ed. (2012). *Constructivist theories of ethnic politics*. New York: Oxford University Press. 500 pp.

Chatterjee, S., A. Mukhopadhyay, and M. Bhattacharyya (2019). "A review of judgment analysis algorithms for crowdsourced opinions". In: *IEEE Transactions on Knowledge and Data Engineering* 32.7, pp. 1234–1248.

Conover, P. J. (1988). "The Role of Social Groups in Political Thinking". In: *British Journal of Political Science* 18.1, pp. 51–76.

Crabtree, C., M. Golder, T. Gschwend, and I. H. Indriason (2019). "It Is Not Only What You Say, It Is Also How You Say It: The Strategic Use of Campaign Sentiment". In: *The Journal of Politics* 82.3, pp. 1044–1060.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805[cs].

Dolinsky, A. O. (2022). "Parties' group appeals across time, countries, and communication channels—examining appeals to social groups via the Parties' Group Appeals Dataset". In: *Party Politics* 0.0.

Dolinsky, A. O., W. Horne, and L. M. Huber (2023). "Parties' group appeals across space and time: an effort towards an automated, large-scale analysis of parties' election manifestos." Working Paper.

Enyedi, Z. (2005). "The role of agency in cleavage formation". In: *European Journal of Political Research* 44.5, pp. 697–720.

Gadjanova, E. (2015). "Measuring parties' ethnic appeals in democracies". In: *Party Politics* 21.2, pp. 309–327.

Gennaro, G. and E. Ash (2022). "Emotion and Reason in Political Language". In: *The Economic Journal* 132.643, pp. 1037–1059.

Goodman, R. and S. Bagg (2022). "Preaching to the Choir? Rhetoric and Identity in a Polarized Age". In: *The Journal of Politics* 84.1, pp. 511–524.

Graf, S., M. Rubin, Y. Assilamehou-Kunz, M. Bianchi, A. Carnaghi, F. Fasoli, E. Finell, M. Gustafsson Sendén, S. E. Shamloo, and J. Tocik (2023). "Migrants, asylum seekers, and refugees: Different labels for immigrants influence attitudes through perceived benefits in nine countries". In: *European Journal of Social Psychology*.

Grimmer, J. and B. M. Stewart (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". In: *Political Analysis* 21.3, pp. 267–297.

He, P., X. Liu, J. Gao, and W. Chen (2021). *DeBERTa: Decoding-enhanced BERT with Disentangled Attention*. arXiv: 2006.03654[cs].

Hersh, E. D. and B. F. Schaffner (2013). "Targeted campaign appeals and the value of ambiguity". In: *The Journal of Politics* 75.2, pp. 520–534.

Ho, J. C.-t. and C.-h. Chan (2023). "Evaluating Transferability in Multilingual Text Analyses". In: *Computational Communication Research* 5.2.

Hobolt, S. B., T. J. Leeper, and J. Tilley (2021). "Divided by the Vote: Affective Polarization in the Wake of the Brexit Referendum". In: *British Journal of Political Science* 51.4, pp. 1476–1493.

Holman, M. R., M. C. Schneider, and K. Pondel (2015). "Gender Targeting in Political Advertisements". In: *Political Research Quarterly* 68.4, pp. 816–829.

Hopkins, D., Y. Lelkes, and S. Wolken (2022). "Which News Goes Viral? Measuring Identity Threats and Engagement with News Media Posts on Twitter and Facebook". In: *PolMeth Europe 2022*. Hamburg.

Horn, A., A. Kevins, C. Jensen, and K. Van Kersbergen (2021). "Political parties and social groups: New perspectives and data on group and policy appeals". In: *Party Politics* 27.5, pp. 983–995.

Howe, P. J., E. Szöcsik, and C. I. Zuber (2022). "Nationalism, Class, and Status: How Nationalists Use Policy Offers and Group Appeals to Attract a New Electorate". In: *Comparative Political Studies* 55.5, pp. 832–868.

Huber, L. M. (2021). "Beyond Policy: The Use of Social Group Appeals in Party Communication". In: *Political Communication* 39.3, pp. 293–310.

Huber, L. M., T. M. Meyer, and M. Wagner (n.d.). "Social group appeals in party rhetoric: Effects on policy support and polarization". In: *The Journal of Politics* ().

Huddy, L. (2001). "From Social to Political Identity: A Critical Examination of Social Identity Theory". In: *Political Psychology* 22.1, pp. 127–156.

Jackson, M. S. (2011). "Priming the Sleeping Giant: The Dynamics of Latino Political Identity and Vote Choice". In: *Political Psychology* 32.4, pp. 691–716.

Jennings, W., S. Bevan, and P. John (2011). "The Agenda of British Government: The Speech from the Throne, 1911-2008". In: *Political Studies* 59.1, pp. 74–98.

Kam, C. D., A. M. N. Archer, and J. G. Geer (2017). "Courting the Women's Vote: The Emotional, Cognitive, and Persuasive Effects of Gender-Based Appeals in Campaign Advertisements". In: *Political Behavior* 39.1, pp. 51–75.

Kitschelt, H. (2000). "Linkages between citizens and politicians in democratic polities". In: *Comparative political studies* 33.6, pp. 845–879.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. 2nd ed. Sage.

Lamont, M. and V. Molnár (2002). "The Study of Boundaries in the Social Sciences". In: *Annual Review of Sociology* 28.1, pp. 167–195.

Lehmann, P., T. Burst, T. Matthieß, S. Regel, A. Volkens, B. Weßels, L. Zehnter, and W. B. F. S. (WZB) (2022). *Manifesto Project Dataset*. Version 2022a.

Licht, H. (2023). "Cross-Lingual Classification of Political Texts Using Multilingual Sentence Embeddings". In: *Political Analysis* 31.3, pp. 366–379.

Lieberman, E. and A. Miller (2021). "Do online newspapers promote or undermine nation-building in divided societies? Evidence from Africa". In: *Nations and Nationalism* 27.1, pp. 238–259.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv: 1907.11692[cs].

Majone, G. (1997). "From the positive to the regulatory state: Causes and consequences of changes in the mode of governance". In: *Journal of public policy* 17.2, pp. 139–167.

Mierke-Zatwarnicki, A. (2023). "Varieties of identity politics: A macro-historical approach". Working Paper.

Miller, A. H., C. Wlezien, and A. Hildreth (1991). "A Reference Group Theory of Partisan Coalitions". In: *The Journal of Politics* 53.4, pp. 1134–1149.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2008). "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict". In: *Political Analysis* 16.4, pp. 372–403.

Muddiman, A., S. C. McGregor, and N. J. Stroud (2019). "(Re)Claiming Our Expertise: Parsing Large Text Corpora With Manually Validated and Organic Dictionaries". In: *Political Communication* 36.2, pp. 214–226.

Nakayama, H., T. Kubo, J. Kamura, Y. Taniguchi, and X. Liang (2018). *doccano: Text Annotation Tool for Human*. Version 1.8.0.

Nteta, T. and B. Schaffner (2013). “Substance and Symbolism: Race, Ethnicity, and Campaign Appeals in the United States”. In: *Political Communication* 30.2, pp. 232–253.

O’Grady, T. (2022). *The transformation of British welfare policy: politics, discourse, and public opinion*. Oxford University Press.

Osnabrügge, M., S. B. Hobolt, and T. Rodon (2021). “Playing to the Gallery: Emotive Rhetoric in Parliaments”. In: *American Political Science Review* 115.3, pp. 885–899.

Pennebaker, J. W., R. L. Boyd, K. Jordan, and K. Blackburn (2015). “The Development and Psychometric Properties of LIWC2015”. In.

Petrocik, J. R. (1996). “Issue Ownership in Presidential Elections, with a 1980 Case Study”. In: *American Journal of Political Science* 40.3, pp. 825–850.

Pitkin, H. F. (1967). *The concept of representation*. University of California Press.

Quinn, K. M., B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev (2010). “How to Analyze Political Attention with Minimal Assumptions and Costs”. In: *American Journal of Political Science* 54.1, pp. 209–228.

Ramshaw, L. and M. Marcus (1995). “Text Chunking using Transformation-Based Learning”. In: *Third Workshop on Very Large Corpora*.

Robison, J., R. Stubager, M. Thau, and J. Tilley (2021). “Does Class-Based Campaigning Work? How Working Class Appeals Attract and Polarize Voters”. In: *Comparative Political Studies* 54.5, pp. 723–752.

Saward, M. (2006). “The representative claim”. In: *Contemporary political theory* 5.3, pp. 297–318.

Sczepanski, R. (2023). “Who are the Cosmopolitans? How Perceived Social Sorting and Social Identities Relate to European and National Identities”. In: *Comparative Political Studies*, p. 00104140231194054.

Simpson, E. and I. Gurevych (2019). “A Bayesian Approach for Sequence Tagging with Crowds”. 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)*. Hong Kong, China: Association for Computational Linguistics. DOI: 10.18653/v1/D19-1101.

Slothuus, R. (2007). “Framing deservingness to win support for welfare state retrenchment”. In: *Scandinavian Political Studies* 30.3, pp. 323–344.

Stückelberger, S. and A. Tresch (2022). “Group Appeals of Parties in Times of Economic and Identity Conflicts and Realignment”. In: *Political Studies* 0.0.

Thau, M. (2019). “How Political Parties Use Group-Based Appeals: Evidence from Britain 1964–2015”. In: *Political Studies* 67.1, pp. 63–82.

Thau, M. (2021). “The Social Divisions of Politics: How Parties' Group-Based Appeals Influence Social Group Differences in Vote Choice”. In: *The Journal of Politics* 83.2, pp. 675–688.

— (2023). “The Group Appeal Strategy: Beyond the Policy Perspective on Party Electoral Success”. In: *Political Studies*, p. 00323217231220127.

Valenzuela, A. A. and M. R. Michelson (2016). “Turnout, status, and identity: Mobilizing Latinos to vote with group appeals”. In: *American Political Science Review* 110.4, pp. 615–630.

Weber, C. and M. Thornton (2012). “Courting Christians: How Political Candidates Prime Religious Considerations in Campaign Ads”. In: *The Journal of Politics* 74.2, pp. 400–413.

White, I. K. (2007). “When Race Matters and When It Doesn't: Racial Group Differences in Response to Racial Cues”. In: *American Political Science Review* 101.2, pp. 339–354.

Wolkenstein, F. (2021). “Revisiting the constructivist turn in political representation”. In: *European Journal of Political Theory* 0.0.

Wolkenstein, F. and C. Wratil (2021). “Multidimensional Representation”. In: *American Journal of Political Science* 65.4, pp. 862–876.

Zollinger, D. (2022). “Cleavage Identities in Voters' Own Words: Harnessing Open-Ended Survey Responses”. In: *American Journal of Political Science* 0.0.

Zuber, C. I., P. J. Howe, and E. Szöcsik (2023). “Policy meets identity: Why and how research on party competition needs to engage with group appeals”.

Online Appendix

"Who are they talking about? Detecting mentions of social groups in political texts with supervised learning"

A Data Descriptives

Our primary dataset records the time series of election manifestos of the two main British parties – the Labour Party and the Conservative Party – from the years 1964 to 2019. We complement this dataset with manifestos for the elections 2015, 2017, and 2019 of the Democratic Unionist Party (DUP), Green Party of England and Wales (Greens), Liberal Democrats (LibDem), Scottish National Party (SNP), and United Kingdom Independence Party (UKIP). We have processed the manifestos in this sample into a sentence-level corpus. We have first collected the raw texts of the relevant manifestos from several sources, including the *Comparative Manifestos Project* (Lehmann et al., 2022) and the *Political Documents Archive* (Benoit et al., 2009). We have then manually processed the raw texts into text files in a way that preserved manifestos' original separation into (sub)chapters, paragraphs, and sentences. The number of sentences in each manifesto are reported in Table A1.

Moreover, we have compiled two additional datasets we use to evaluate the generalizability and transferability of our approach (see Appendix E.3). The first additional dataset records German parties' manifestos. It comprises the Christian Democratic Union (CDU) and the Social Democratic Party of Germany (SPD) election manifestos for the six elections from 2002 to 2021, and the manifestos of the Alliance'90/Greens (Greens), The Left (LINKE), the Free Democratic Party (FDP), and the Alternative for Germany (AfD) for the elections 2013, 2017, and 2021. We have again processed the raw texts of these manifestos into a sentence-level corpus and the number of sentences in each manifesto are reported in Table A2.

Our second additional dataset records British parties' question time speeches in the UK *House of Commons*. It includes all parties covered in our British manifestos corpus and covers the period 2013 to 2019. We have again process documents in this sample into a sentence-level corpus. We have first extracted the relevant speeches from the ParlSpeech2 dataset (Rauh and Schwalbach, 2020). We have then identified all question time agenda items between 2013 and 2019, subset the corpus to speeches delivered by the parties on our case selection, cleaned encoding errors from the speech texts, and segmented speech texts into sentences. The number of sentences for each party–year are reported in Table A3.

Table A1. Number of sentences in UK party manifestos.

Party	1964	1966	1970	1974-02	1974-10	1979	1983	1987	1992	1997	2001	2005	2010	2015	2017	2019
Conservatives	399	276	550	587	677	425	634	1011	1614	1136	724	395	1357	1232	1152	849
DUP														100	273	345
Greens														1639	130	828
Labour	408	501	539	198	419	481	1184	456	662	986	1598	1112	1297	911	1024	1071
LibDem														1256	804	991
SNP														616	633	689
UKIP														928	837	

Table A2. Number of sentences in German party manifestos.

Party	2002	2005	2009	2013	2017	2021
Alliance'90/Greens				4993	3802	3674
Alternative for Germany				71	969	1394
Christian Democratic Union	1219	714	1743	2509	1292	2579
Free Democratic Party				2357	2020	2073
Social Democratic Party of Germany	1474	852	2003	2493	2330	1443
The Left				2271	3746	4488

Table A3. Number of sentences in UK House of Commons data.

Party	2013	2014	2015	2016	2017	2018	2019
Con	33593	35089	33614	40571	36254	44538	38643
DUP	289	316	278	446	342	494	401
Lab	12101	12797	9842	10363	10216	13499	11866
LibDem	6150	6494	2176	501	591	816	843
SNP	284	244	1468	2970	2445	3010	2820

B Annotation Details

Table B4. Number of sentences distributed for annotation per annotation job.

Annotation job	N
UK: Labour and Conservative manifestos (1964-2015)	6334
UK: Labour and Conservative manifestos (2017 and 2019)	900
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	1362
UK: House of Commons speeches (2013-2019)	1575
Germany: CDU and SPD manifestos (2002-2021)	1500
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	1427

Table B5. Number of sentences annotated by two or one coder per dataset.

	N coders	
Annotation job	2	1
UK: Labour and Conservative manifestos (1964-2019)	2399	4835
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	588	774
UK: House of Commons speeches (2013-2019)	524	1050
Germany: CDU and SPD manifestos (2002-2021)	299	1201
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	300	1127

Table B6. Categories in our group mention detection coding scheme.

Category	Code	Description
Social group	`SG`	Collectives of people with common characteristics. Examples are *families*, *young people*, *households in the poorest third of the population*, or *the rich*.
Political group	`PG`	Collectives of people that act on behalf of or are associated (i.e. by other actors) with a political party. Examples are *the Conservatives*, *Liberal MPs*, or *the Labour Government*.
Political institution	`PI`	All institutions belonging to the executive, legislative, or judicatory on the inter-, national, or sub-national level. For example, this includes *the European Union*, *NATO*, *the government*, *the police*, and *local authorities*.
Organization, public institution, or collective actor	`ORG`	Organizations and associations that do not have direct political representation in parliament and government. Examples are *firms*, *(trade) unions*, *schools*, *hospitals*, *public corporations*, or *the Black Lives Matter movement*.
Implicit social group reference	`ISG`	Text passages that refer to a group but only explicitly. Examples are *communities*, *everyone in Britain*, *the public*, *the nation*, etc.
"unsure"		This miscellaneous category can be applied if a coder identified a potential group mention but they were not sure how to categorize it.

Table B7. Proportions of sentences by annotation job that contain (i) no mention at all, (ii) no social group mention (but min. one mention of another group type), or (iii) at least one social group mention.

Annotation job	no mention at all	no social group mention	min. one social group mention
UK: Labour and Conservative manifestos (1964-2019)	2146 (0.30)	3054 (0.42)	2034 (0.28)
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	350 (0.26)	627 (0.46)	385 (0.28)
UK: House of Commons speeches (2013-2019)	464 (0.29)	703 (0.45)	407 (0.26)
Germany: CDU and SPD manifestos (2002-2021)	731 (0.49)	455 (0.30)	314 (0.21)
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	705 (0.49)	443 (0.31)	279 (0.20)

Table B8. Summary statistics of sentence-level inter-coder agreement scores by annotation job in doubly annotated sentences that are coded by at least one coder as containing at least one group mention annotation. The top panel reports agreement when counting only group mention annotations (and treating all other annotations as outside a span). The bottom panel, in contrast, reports agreement when considering all five group categories in our coding scheme. *Note:* Sentence with no annotation by either coder omitted because agreement is 100% in all.

		All group categories			Social group vs. none (binary)		
Annotation job	N	10% ptl.	Mean	Median	10% ptl.	Mean	Median
min. one social group mention							
UK: Labour and Conservative manifestos (1964-2019)	724	0.88	0.91	0.96	0.92	0.94	1.00
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	177	0.82	0.89	0.96	0.91	0.92	1.00
UK: House of Commons speeches (2013-2019)	144	0.84	0.89	0.95	0.86	0.91	0.96
Germany: CDU and SPD manifestos (2002-2021)	61	0.92	0.93	1.00	0.93	0.95	1.00
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	62	0.88	0.92	0.95	0.92	0.95	1.00
no social group mention							
UK: Labour and Conservative manifestos (1964-2019)	1022	0.88	0.92	0.95	1.00	1.00	1.00
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	264	0.89	0.92	0.96	1.00	1.00	1.00
UK: House of Commons speeches (2013-2019)	247	0.89	0.93	1.00	1.00	1.00	1.00
Germany: CDU and SPD manifestos (2002-2021)	98	0.84	0.88	0.92	1.00	1.00	1.00
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	103	0.83	0.88	0.93	1.00	1.00	1.00

Table B9. Distribution of disagreement patterns in sentences segments with intercoder disagreement by annotation job.

		Agreement				
Annotation job	N	none	on first part	on mid part	on last part	other pattern
UK: Labour and Conservative manifestos (1964-2019)	329	0.559	0.088	0.012	0.264	0.076
UK: DUP, Greens, LibDem, SNP, and UKIP (2015-2019)	93	0.613	0.140	0.032	0.161	0.054
UK: House of Commons speeches (2013-2019)	92	0.522	0.163	0.022	0.283	0.011
Germany: CDU and SPD manifestos (2002-2021)	24	0.750		0.042	0.208	
Germany: AfD, B90/GRÜNE, FDP, and LINKE (2013-2021)	29	0.759	0.069		0.172	

B.1 Annotation aggregation

Because we have collected annotations from two coders for some sentences, we need to aggregate these annotations into a single set of word-level labels per sentence. We follow the rich computer science literature on annotation aggregation (cf. Chatterjee et al., 2019) and fit a Bayesian sequence combination model (Simpson and Gurevych, 2019). This method estimates the "true" word-level labels from multiple annotations per sentence while accounting for coders' (estimated) annotation abilities. In addition, it accounts for serial dependencies in words' labels, which is important since the words in a sentence are not independent of each other.

To aid the model in identifying the true sequence labels from our coders' annotations, we have reviewed the disagreements in the 612 British doubly-annotated manifesto sentences with 90% or less word-level agreement. We have then used our expert annotations as informative priors when fitting the Bayesian model.

Table B10. Descriptive statistics of group mentions in labeled sentences in our UK party manifesto corpus.

		Mentions		N tokens				
Category	Share any mention	N	N_{unique}	Mean	25% perc.	50% perc.	75% perc.	Max.
social group	0.306	3373	1978	3.054	1	2	4	28
political group	0.151	1375	208	1.794	1	2	2	31
political institution	0.203	2125	913	2.397	1	2	3	26
collective actor	0.193	2118	1135	2.275	1	2	3	14
implicit group reference	0.108	998	255	1.496	1	1	2	5

Table B11. Descriptive statistics of group mentions in labeled sentences in our German party manifesto corpus.

		Mentions		N tokens				
Category	Share any mention	N	N_{unique}	Mean	25% perc.	50% perc.	75% perc.	Max.
social group	0.272	1028	649	2.260	1	1	3	17
political group	0.061	200	62	1.850	1	2	2	10
political institution	0.111	391	215	1.831	1	1	2	17
collective actor	0.136	499	383	1.623	1	1	2	7

C Additional classifier results

Table C12. Summary of test set performances in terms of the F1 score of our group mention detection classifier trained and evaluated on our corpus of labeled UK manifesto sentences. Values (in brackets) report the average (90% quantile range) of performances of 25 different classifiers trained in a 5-times repeated 5-fold cross-validation scheme. Rows report results for the different group categeries included in our coding scheme. Columns distinguish between different evaluation schemes (i.e., different ways to compute the F1 score). *Note:* `seqeval` is the strict metric proposed by Ramshaw and Marcus (1995) and implemented by Nakayama (2018).

Category	`seqeval`	Span level	Sentence level
SG	0.87 [0.80, 0.93]	0.96 [0.94, 0.99]	0.97 [0.94, 0.99]
PG	0.91 [0.85, 0.95]	0.98 [0.96, 0.99]	0.99 [0.99, 1.00]
PI	0.85 [0.77, 0.91]	0.96 [0.94, 0.98]	0.98 [0.96, 0.99]
ORG	0.83 [0.72, 0.92]	0.96 [0.93, 0.98]	0.98 [0.97, 0.99]
ISG	0.80 [0.66, 0.92]	0.97 [0.94, 0.99]	0.97 [0.95, 0.99]

Table C13. Average span-level social group mention recall in test set of classifier trained and evaluated on corpus of labeled UK manifesto sentences by number of words in span. *Notes:* Test-set spans grouped into bins based on the number of words they contain at the 50, 90, and 97.5% percentile values. For example, bin (0, 2] contains social group mentions that span up to two words, bin (2, 6] mentions that span between three to six words, etc. Recall values in square brackets report the 90% confidence intervall computed from bootstrapped values.

Words in span	Average span-level recall	N_{span}
(0, 3]	0.88 [0.84, 0.92]	222
(3, 6]	0.83 [0.77, 0.88]	69
(6, 9]	0.71 [0.52, 0.86]	17
(9, 17]	0.67 [0.48, 0.87]	15

Table C14. Most frequent words group-based appeals manually identified by Thau (2019) our classifier has not predicted to belong to a social group mention. One row per type of group according to Thau's categorization, focusing on those where our classifier perfroms relatively poorly in terms of the social group recall. *Note:* Values in parentheses indicate the number of occurences in the relevant subset of Thau's data.

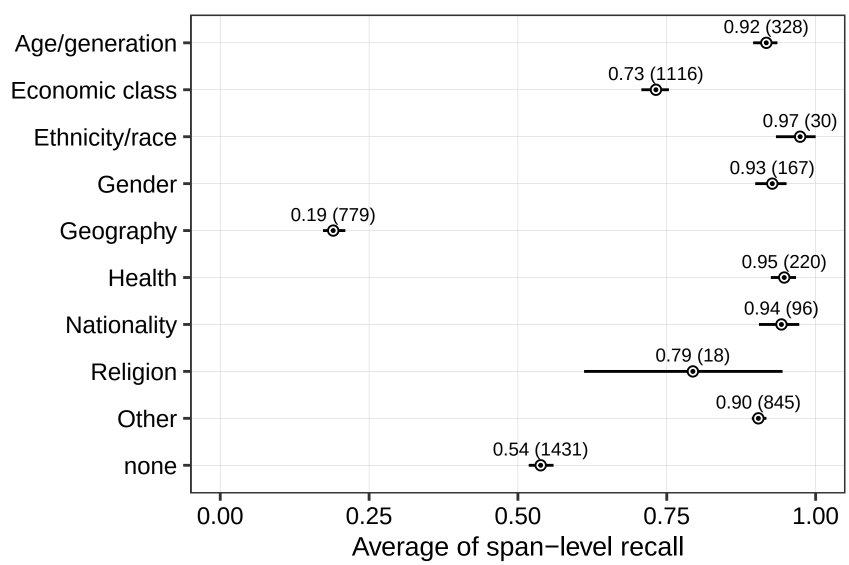
Thau's categorization	Word (N)
Economic class	businesses (51), companies (49), small (42), firms (36), management (18), british (14), business (11), large (8), banks (7), energy (7), local (7), producers (7)
Geography	areas (58), cities (54), local (45), towns (45), wales (37), scotland (34), communities (32), regions (30), england (27), country (26)
Religion	churches (2), faith (2), communities (1), groups (1)
none	people (68), country (54), unions (53), trade (43), everyone (27), community (24), individuals (22), ' (21), officers (20), public (20), union (20)

D Comparison to data collected by Thau (2019)

Figure D1, shows that a social group detection classifier trained with our appraoch performs overall consistently across the group categories distinguished by Thau (2019). It achieves average mention-level recall values above 0.90 in most categories.

The three exceptions to this pattern are explained by how our coding instructions diverge from Thau's. Many of the mentions included in Thau's "Economic class" and "Religion" group categories are considered as coded as ORG mentions according to our coding scheme. Similarly, many of the mentions included in Thau's "Geography" category are references to geo-spatial entities (e.g., Scotland), which we have explicitly excluded unless they were part of an implicit social group reference.

Figure D1. Average mention-level recall of predicted social group mentions in group-based appeals manually identified by Thau (2019). Recall computed by assuming that group mentions identified by Thau (2019) are "true" *social* group mentions and comparing them to token-level labels predicted by our group mention detection classifier trained on labeled sentences from all UK party manifestos in our sample. The x-axis indicates the average share of tokens in a "true" mention the classifier has predicted correctly (values with number of spans plotted above points for readability). The y-axis indicates the type of group according to Thau's categorization. Horizontal lines report the 90% confidence intervall computed from bootstrapped recall values.



E Additional analyses: Resource effectiveness, generalization, and transferability

The results presented in our paper demonstrate the high reliability, validity, and flexibility of our supervised group mentions detection approach. However, applied researchers might want to adopt our approach to study group-based rhetoric in other domains or languages and potentially with a different target concept. The practical utility of our approach in other research settings thus stands and falls with its resource intensiveness and ability to generalize to unseen data, domains, and languages. This section summarizes the results of additional analyses that shed light on these questions

E.1 Resource effectiveness

We first address the question of how the number of labeled sentences used to fine-tune a group mention detection classifier affects its reliability (cf. Barberá et al., 2021; Licht, 2023). Considering that the results we present in Table 2 are based on classifiers trained on 6,000 to 7,000 labeled sentences, researchers likely wonder how our approach fares with less labeled data.

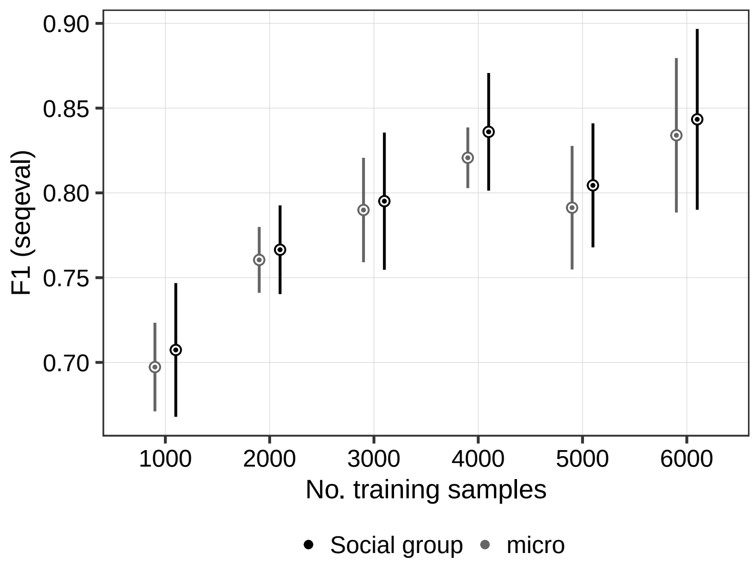
We present evidence on this question in Figure E2. It shows that performances similar to those reported in Table 2 can be achieved with far fewer labeled sentences. For example, the average `seqeval` F1 score for the social group category is already 0.80 when training on only 3,000 labeled sentences, and the 90% confidence interval overlaps with those of the average performance achieved when training on 6,000 labeled sentences. In addition, we show further below that a classifier trained on some source documents (e.g., UK manifestos) can be reliably adapted to a target domain (e.g., German manifestos) with relatively few additional labeled sentences. This suggests that in practice, researchers can start with our pre-trained classifiers and adapt them to their target domain and use cases.

E.2 Generalization

We next turn to the question of generalization. A first condition for generalization is that the high predictive performance we have documented above is not simply due to classifiers' ability to "remember" mentions it has already "seen" during training. We scrutinize this possibility by comparing the performance of a classifier trained on 6,861 labeled sentences sampled from all UK parties' manifestos in correctly predicting the labels of group mentions in the test set that did not occur in the training set.[17] The average span-level recall for the

[17]Specifically, we have identified the unique set of social group mentions contained in the training data and created an indicator for mentions in the test set that is `1` if it was in the training data ("seen") and `0` otherwise ("unseen").

Figure E2. Summary of test set performances as function of training data size of our group mention detection model trained and evaluated on corpus of labeled UK manifesto sentences. Points (line ranges) report the average (± 1 std. dev.) of performances of 5 different classifiers trained with different random seeds. The y-axis indicates the `seqeval` F1 score achieved. The x-axis indicates the number of sentences in the training set. Colors distinguish between the micro performance and the social group mention category-specific performances. *Note:* `seqeval` is the strict metric proposed by Ramshaw and Marcus (1995) and implemented by Nakayama (2018).



social group mentions "seen" during training is 0.89; for "unseen" social group mentions, it is 0.82. This indicates that while not free of reliability losses, generalization into unseen data is possible with our approach.[18]

A further potential threat to the generalization of our findings — and thus the utility of our proposed method — is that all our annotations were collected applying a single social group definition. While we have oriented our conceptualization at the existing literature, applied researchers' target concept might deviate from ours. We believe that our proposed approach is still useful for these researchers, given that classifiers trained on UK party manifestos perform reliably in each of the five group categories included in our coding scheme. For example, the average `seqeval` F1 scores for detecting mentions of political groups and political institutions are 0.91 and 0.81, respectively (see Table C12 for details). This suggests that similar levels of reliability as we report can be achieved with deviating social group conceptualizations and coding instructions.

[18] An example of an "unseen" mention where our classifier showed 0 recall is 'the Next Generation.' An example of an "unseen" mention where our classifier showed perfect recall is 'every school leaver that gets the grades.'

E.3 Transferability

A final way to assess the generalization potential of our approach is to test whether it performs well when applied to labeled data from different domains, languages, and contexts. We evaluate the capabilities of our proposed method in this regard by assessing its performance in zero- and few-shot transfer (cf. Licht, 2023; Osnabrügge et al., 2021). Zero-shot transfer means to classify sentences from documents in a "target domain" (e.g. parliamentary speech) using a classifier solely trained on sentences from a "source domain" (e.g., party manifestos). In turn, few-shot transfer means classifying sentences from target-domain documents using a classifier that has mainly been trained on source-domain sentences *and* also a small portion of target-domain documents. For example, these two types of transfer mirror the practical research settings in which a researcher applies our classifier trained on UK party manifestos to label sentences from German parties' manifestos or UK parliamentary speeches without (zero-shot) or with (few-shot) adding labeled sentences from their target corpus in the training data.

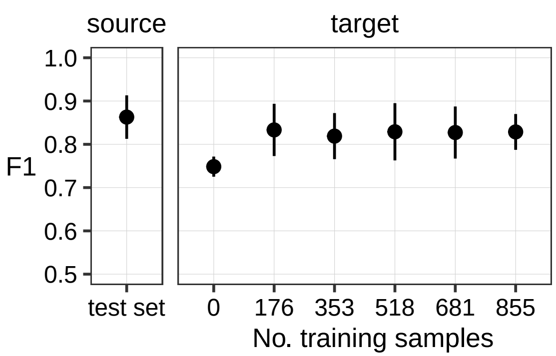
We have run extensive transfer experiments to probe the performance of our approach in zero- and few-shot transfer. In a first "cross-party transfer" experiment, we use the manifestos of the smaller British parties in our corpus (DUP, Greens, SNP, and UKIP) as target documents and the Conservatives and Labour Party manifestos as source documents. In a second experiment, we examine "cross-lingual transfer" and use German parties' manifestos as target documents and British parties' (English-language) manifestos as source documents (cf. Licht, 2023). In a third experiment, we examine "cross-domain transfer" and use sentences sampled from British *House of Commons* speeches as target documents and British parties' manifestos as source documents (cf. Osnabrügge et al., 2021). The datasets are for these experiments are described in Appendices A and B.

In each of these experiments, we have started with a classifier trained only on labeled sentences from the source domain and evaluated this classifier on test sets sampled from the source and target domains, respectively. This allows us to compare the zero-shot transfer performance to the baseline of no transfer. We have then incrementally used portions of labeled sentences in the target-domain 50% training split to adapt the classifier to the target domain through continued training. For each experiment, we have repeated this process with five different seeds to account for uncertainty in fine-tuned classifiers' performances.

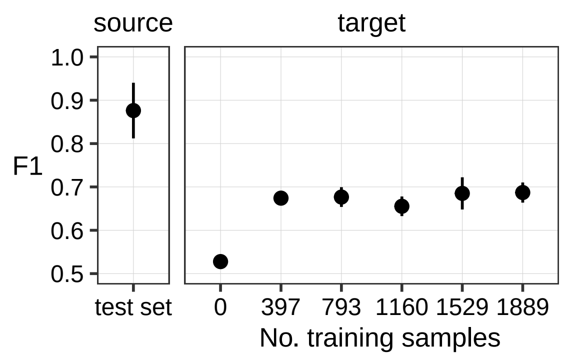
The results from these experiments are reported in Figure E3. The first data points at x-axis values of 0 in the right-hand plot panels report the results for zero-shot transfer. This shows that zero-shot transfer classification comes with reliability losses – especially when this involves transfer across party systems and language or institutional settings. However, at least in the cases of cross-party and cross-domain transfer, the reliability losses are relatively modest.

Yet, Figure E3 also shows that the reliability of transfer to the target domain can be

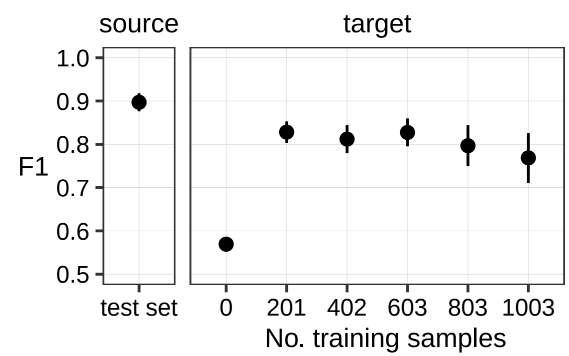
Figure E3. Summary of test set performances in cross-party, cross-country, and cross-domain transfer, respectively. The y-axis indicates the performance of classifiers trained on annotated manifesto sentences from the source domain (e.g., British manifestos) when evaluated on sentences from the target domain (e.g., German manifestos) in terms of the `seqeval` F1 score. Points (line ranges) report the average (± 1 std. dev.) of performances of 5 different classifiers trained with different random seeds. For the "target" panel, the x-axis reports whether the classifier trained on source-domain documents was evaluated on the target-domain documents without adapting it to the target texts ("zero shot") or, if not, how many target-domain sentences were used to adapt the classifier through continued training. For comparison, we report the performance in a held-out set of source-domain sentences in the "source" panel.



(a) Cross-party transfer: From UK Labour and Conservative party manifestos ("source") to other parties' manifestos ("target").



(b) Cross-country transfer: From British parties' manifestos ("source") to German manifestos ("target").



(c) Cross-domain transfer: From British parties' manifestos ("source") to their speeches in the *House of Commons* ("target").

improved through few-shot learning, that is, continuing to train the classifier pre-trained on source-domain documents with a few labeled sentences from the target domain. In all three transfer experiments, compared to the zero-shot baseline, classifiers' reliability increases when continuing to train the source-domain classifier with a few hundred labeled sentences. As a point in case, in the cross-party transfer experiment (Figure E4a), continuing to train it with only 176 labeled sentences (10% of the target corpus) allows matching the performance achieved in the source-domain test set. Yet, continuing to train with more labeled data from the target domain does not improve classification performance in the target domain further. The results for cross-lingual transfer are not as strong (see Figure E4b), which is likely explained by the fact that in this setup, we need not only to transfer across languages but also party systems and political cultures. Nevertheless, even in the few-shot cross-lingual transfer experiment 10% of the labeled target corpus (397 labeled sentences) already yield substantial performance improvements relative to the zero-shot baseline. We find a similar initial improvement for cross-domain transfer from UK manifestos to parliamentary speech (see Figure E4c). However, as we continue to adapt the source-domain classifier with more and more labeled parliamentary speech sentences, classifiers' target-domain performance becomes more uncertain and slightly decreases as a consequence.

Overall, our findings on the transferability of our approach suggest that, in practice, researchers can start off with our pre-trained classifiers and adapt them to their target domain and use cases. We thus believe that our approach enables even less well-endowed researchers to size the scalability advantage of our proposed approach. Further, our results suggest that by fine-tuning on a small but diverse and potentially multilingual set of labeled sentences from different domains or countries, our approach can enable reliable detection and retrieval of (seen *and* unseen) social group mentions in political texts. Our results thus highlight our approach's great promises for large-scale comparative research.

F Additional results

F.1 Relation between emotional language and group mentions

Table F15. Logistic regression coefficient estimates.

	Emotions	Positive emotions	Negative emotions
Social group mention(s)	0.198^{***}	0.115^{**}	0.153^{**}
	(0.034)	(0.036)	(0.059)
progressive–conservative position	0.007^{***}	0.006	-0.002
	(0.002)	(0.003)	(0.005)
state–market position	0.002	0.002	0.007^{*}
	(0.001)	(0.002)	(0.003)
Prime minister party	-0.082^{*}	0.064	-0.319^{***}
	(0.038)	(0.045)	(0.048)
N tokens	0.068^{***}	0.061^{***}	0.035^{***}
	(0.002)	(0.002)	(0.002)
AIC	31802.674	33591.228	26563.127
BIC	31974.047	33762.602	26734.501
Log Likelihood	-15880.337	-16774.614	-13260.563
Deviance	31760.674	33549.228	26521.127
Num. obs.	25865	25865	25865

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
All models include election fixed effects.
Standard errors clustered by party and election.

Table F16. Logistic regression coefficient estimates from Regressing binary indicator of use of (positive/negative/both) emotion words in sentence on mention of at least one social group in the sentence, excluding cases where *all* emotion words in a sentence with one or more group mentions belong to the group mention(s).

	Emotions	Positive emotions	Negative emotions
Social group mention(s)	0.078^{*}	0.146^{***}	-0.005
	(0.034)	(0.038)	(0.065)
progressive–conservative position	0.007^{***}	0.006^{*}	-0.002
	(0.002)	(0.003)	(0.004)
state–market position	0.002	0.002	0.008^{*}
	(0.001)	(0.002)	(0.003)
Prime minister party	-0.087^{*}	0.063	-0.335^{***}
	(0.040)	(0.044)	(0.049)
N tokens	0.069^{***}	0.063^{***}	0.037^{***}
	(0.002)	(0.002)	(0.002)
AIC	31355.291	32775.498	25453.862
BIC	31526.224	32946.430	25624.794
Log Likelihood	-15656.646	-16366.749	-12705.931
Deviance	31313.291	32733.498	25411.862
Num. obs.	25327	25327	25327

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
All models include election fixed effects.
Standard errors clustered by party and election.

Table F17. Logistic regression coefficient estimates from regressing binary indicator of (positive/negative/both) emotion word use in sentences on mention of at least one social group in the sentence in all party manifestos in our UK corpus.

	Emotions	Positive emotions	Negative emotions
Social group mention(s)	0.211^{***}	0.126^{***}	0.251^{***}
	(0.031)	(0.030)	(0.048)
Progressive–conservative position	-0.004	-0.008^{*}	0.003
	(0.002)	(0.004)	(0.002)
State–market position	0.009^{***}	0.011^{***}	0.003
	(0.002)	(0.003)	(0.003)
Prime minister party	-0.051	0.102^{*}	-0.302^{***}
	(0.038)	(0.051)	(0.047)
N tokens	0.064^{***}	0.055^{***}	0.029^{***}
	(0.002)	(0.002)	(0.002)
AIC	43479.651	46588.227	37673.866
BIC	43657.929	46766.505	37852.144
Log Likelihood	-21718.825	-23273.114	-18815.933
Deviance	43437.651	46546.227	37631.866
Num. obs.	35934	35934	35934

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
All models include election fixed effects.
Standard errors clustered by party and election.

Table F18. Logistic regression coefficient estimates from Regressing binary indicator of use of (positive/negative/both) emotion words in sentence on mention of at least one social group in the sentence in all party manfistos from elections 2015 onwards in our UK corpus.

	Emotions	Positive emotions	Negative emotions
Social group mention(s)	0.229^{***}	0.104^{**}	0.451^{***}
	(0.043)	(0.040)	(0.055)
Progressive–conservative position	-0.010^{***}	-0.015^{***}	0.006^{**}
	(0.002)	(0.003)	(0.002)
State–market position	0.027^{***}	0.037^{***}	-0.010
	(0.006)	(0.008)	(0.006)
Prime minister party	-0.113	0.069	-0.319^{***}
	(0.068)	(0.096)	(0.063)
N tokens	0.059^{***}	0.052^{***}	0.022^{***}
	(0.003)	(0.003)	(0.002)
AIC	19049.863	20883.231	17578.806
BIC	19111.459	20944.826	17640.401
Log Likelihood	-9516.932	-10433.615	-8781.403
Deviance	19033.863	20867.231	17562.806
Num. obs.	16308	16308	16308

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$.
All models include election fixed effects.
Standard errors clustered by party and election.

References

Barberá, P., A. E. Boydstun, S. Linn, R. McMahon, and J. Nagler (2021). "Automated Text Classification of News Articles: A Practical Guide". In: *Political Analysis* 29.1, pp. 19–42.

Benoit, K., T. Bräuninger, and M. Debus (2009). "Challenges for Estimating Policy Preferences: Announcing an Open Access Archive of Political Documents". In: *German Politics* 18.3, pp. 441–454.

Chatterjee, S., A. Mukhopadhyay, and M. Bhattacharyya (2019). "A review of judgment analysis algorithms for crowdsourced opinions". In: *IEEE Transactions on Knowledge and Data Engineering* 32.7, pp. 1234–1248.

Lehmann, P., T. Burst, T. Matthieß, S. Regel, A. Volkens, B. Weßels, L. Zehnter, and W. B. F. S. (WZB) (2022). *Manifesto Project Dataset*. Version 2022a.

Licht, H. (2023). "Cross-Lingual Classification of Political Texts Using Multilingual Sentence Embeddings". In: *Political Analysis* 31.3, pp. 366–379.

Nakayama, H. (2018). *seqeval: A Python framework for sequence labeling evaluation*. Version 1.2.2.

Osnabrügge, M., E. Ash, and M. Morelli (2021). "Cross-Domain Topic Classification for Political Texts". In: *Political Analysis* First view, pp. 1–22.

Ramshaw, L. and M. Marcus (1995). "Text Chunking using Transformation-Based Learning". In: *Third Workshop on Very Large Corpora*.

Rauh, C. and J. Schwalbach (2020). *The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies*.

Simpson, E. and I. Gurevych (2019). "A Bayesian Approach for Sequence Tagging with Crowds". 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)*. Hong Kong, China: Association for Computational Linguistics. DOI: 10.18653/v1/D19-1101.

Thau, M. (2019). "How Political Parties Use Group-Based Appeals: Evidence from Britain 1964–2015". In: *Political Studies* 67.1, pp. 63–82.