# Swimming with the Tide?
# Positional Claim Detection across Political Text Types

**Nico Blokker**[1], **Erenay Dayanik**[2], **Gabriella Lapesa**[2], and **Sebastian Padó**[2]

[1]SOCIUM, University of Bremen, Germany
[2]IMS, University of Stuttgart, Germany
{*blokker*}*@uni-bremen.de*
{*erenay.dayanik, gabriella.lapesa, sebastian.pado*}*@ims.uni-stuttgart.de*

## Abstract

Manifestos are official documents of political parties, providing a comprehensive topical overview of the electoral programs. Voters, however, seldom read them and often prefer other channels, such as newspaper articles, to understand the party positions on various policy issues. The natural question to ask is how compatible these two formats (manifesto and newspaper reports) are in their representation of party positioning. We address this question with an approach that combines political science (manual annotation and analysis) and natural language processing (supervised claim identification) in a cross-text type setting: we train a classifier on annotated newspaper data and test its performance on manifestos. Our findings show a) strong performance for supervised classification even across text types and b) a substantive overlap between the two formats in terms of party positioning, with differences regarding the salience of specific issues.

## 1 Introduction

Electoral programs described in party manifestos are official documents of political parties. Not only do they capture their positions on a broad variety of policy issues, but they are also authoritative for the entire party. In this sense, they are fundamentally different from statements issued only by factions or solitary members of a party (Budge et al., 2001, p. 216). Ordinary citizens, however, rarely read the manifestos (Budge, 1987; Volkens and Bara, 2013). Arguably, an informed voter is more likely to be exposed to electoral programs indirectly through another medium, namely the newspaper. News articles present the reader with the *political claims* put forward by different parties and often directly contrast them. More specifically, we adopt the definition of claims as any form of politically motivated demand or action (both verbal and non-verbal) of deliberate actors (Koopmans and Statham, 1999).
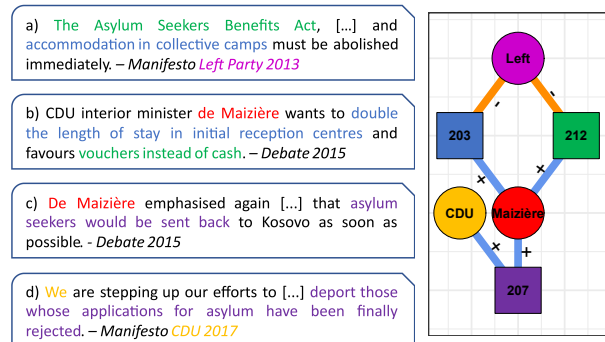


Figure 1: Mapping cross-genre policy proposals

Information about claims is routinely used by readers to infer the parties' standpoints. This places parties in a delicate situation: They largely rely on newspapers and other media to convey their claims to the voter to gain their approval in elections (Robertson, 1976).

Recent advances in the application of NLP methods to political reporting can reliably extract such claims with models trained on manually annotated claims in newspapers (Dayanik and Padó, 2020; Haunss et al., 2020). From a NLP perspective, this gives rise to the following question which we will address in this paper: Can we transfer existing models for claim identification to party manifestos without substantial loss of performance? If so, this would enable political scientists to address substantive questions such as: To what extent do manifestos of specific parties and the positions of their members as reported in newspaper articles overlap? How do differences articulate themselves in practice?

One way to address these questions is to directly compare the agreement between claims from the two formats[1]. Figure 1 (left panel) shows four sentences containing political claims on the pol-

---

[1]We use *format* and *genre* interchangeably to describe *text types*.

icy issue of migration. While the first sentence is translated from the party manifesto of the German left-wing party *Linke* (Left) from the federal election 2013, the last one is from the 2017 manifesto of the conservative *Christian Democratic Union* (CDU). Both sentences in the middle are statements from the German minister of the interior (CDU) published in newspaper articles in 2015.

Despite the discrepancy in time and genre, a) and b) as well as c) and d) encode the same political claims: In b) the minister supports (+) the idea to give refugees vouchers instead of cash (claim category 212[2]), and he proposes to extend the length of central accommodation of refugees (203). In a) both claims are opposed (-) by the Left party (benefits in kind are regulated by the 'Asylum Seekers Benefits Act' in Germany). Analogously, in c) and d) the minister and his party CDU both express support to deport asylum seekers (claim 207).

On a substantive level, the sentences are very similar. Yet linguistically, they are quite different. While manifestos are written in direct speech and express the will of the party, the quotes from the public (newspaper) debate hold indirect speech that can be attributed to an individual actor. Besides, they often differ in the lexical choices, e.g. colloquial vs. technical terms, as in "sent back" vs. "deport" in Figure 1 c) and d)). To exacerbate matters, differences exist within genres in addition to the variability between them.

The purpose of this paper is to determine whether computational models for claim detection in newspaper articles (Dayanik and Padó, 2020) can be applied directly to party manifestos. Our findings indicate that there is indeed a strong transferability and a substantive overlap between estimated party positions between the text types on the issue of migration. These findings are limited in scope to the German case and require further validation given the heterogeneity of both parties and policy issues.

Nevertheless, the contributions of our work address several points. First, we take a further step in the strive for scaling up the semi-automatic identification of tangible policy instruments in political texts (Haunss et al., 2020) by introducing cross-

text type evaluation. Second, at a theoretical level, we provide further experimental insights into the relationship between party manifestos and newspaper debates: we analyse the substantive overlap of both formats considering the underlying claim-distribution per party in the two most recent federal elections in Germany (2013 and 2017) and the interim public newspaper discourse (2015) to see whether previous findings from the political science literature (cf. Section 2) hold. Third, we extend the scope of our investigation to a concrete application scenario, namely a discourse network analysis of the debate at issue (right panel, Figure 1). In this relational approach, the actor-claim-dyad is the first building block of a more complex, bipartite network structure. By abstracting from the actual text it is possible to directly contrast, compare, and even combine the two different formats on a network level. In this perspective, individual actors/parties and claims are two distinct types of vertices that are connected via edges that express support or opposition (Leifeld, 2009, 2016). We use this to demonstrate how our approach translates from newspaper data to party manifestos and how it leads to deeper conceptual insights.

Section 2 illustrates how our approach relates to the literature and compares characteristics of the distinct data sources. Section 3 examines existing modeling approaches to claim identification. The results are presented in Section 4 and 5 and their relevance is discussed in Section 6.

## 2   Related Work: Debate and Manifestos

**Political Science.**   The inter-linkage of public (newspaper) debates and party documents is currently widely investigated (Schwarzbözl et al., 2019; Haselmayer et al., 2019; Merz, 2018). Political parties rely on mass media to distribute their declarations of intent to the voter (Robertson, 1976; Bara, 2006), who in turn holds the party responsible for the promises made (American Political Science Association, 1950; Thomassen, 1994; Adams, 2001, for a critical discussion see Mair, 2009). Given these assumptions, it is natural to infer a substantial overlap between the content of domain-specific newspaper articles and the corresponding sections in party manifestos. Therefore, the identification of proposed policy instruments – *political claims*, or, in the electoral context, *pledges* (Rallings, 1987), in one of the text types arguably parallels the existence in the other.

---

[2]We use the claim categories and codes for the migration debate proposed by Lapesa et al. (2020). The data set can be found here: `hdl.handle.net/11022/1007-0000-0007-DB07-B` and the coding scheme here: `https://github.com/mardy-spp/mardy_acl2019/blob/master/codebook.pdf`

Despite the mediating role played by newspapers in the diffusion process of disseminating party visions, vast differences between the text types exist on a conceptual level. While electoral programs paint a cohesive and unified picture of the party line, media coverage highlights conflict within parties. Parties do not control the content written in newspaper articles. Instead, their message is filtered and interpreted by editors and authors. In the news, parties compete with each other, trying to establish themselves and appeal to the reader (Helbling and Tresch, 2011; Green-Pedersen and Mortensen, 2015).

Hence, existing studies measure and validate the substantive agreement between several (media) genres both in respect to party positioning on certain policy issues as well as the amount of attention (salience) these issues receive (Ray, 2007; Netjes and Binnema, 2007). For instance, comparisons are carried out between expert survey and party manifestos (Benoit and Laver, 2007; Marks et al., 2007) and, more recently, with the addition of newspaper articles as a third data source. Helbling and Tresch (2011) find that while party positions are mostly congruent between manifestos and media coverage in the same election campaign, they differ when it comes to the salience of specific issues. Thus, the media bias appears to have a stronger impact on the selection of certain topics than on the content's accuracy regarding party positions (Helbling and Tresch, 2011, p. 180). One natural characteristic is that the substantive density (e.g. as a ratio of claims to text) is much higher in electoral programs than in newspaper articles.

**Natural Language Processing.** The literature concerning NLP support corpus-based investigations in political (and more broadly social) science is quite heterogeneous. The first group of approaches targets the facilitation of the annotation procedure (traditionally carried out by hand) with argument mining or machine learning. The goal is to speed up annotation without losing in quality (Cabrio and Villata, 2018; Lippi and Torroni, 2015, 2016). The second line of research targets the direct automatic analysis of politic debates in textual form. As far as political claim analysis in newspaper articles is concerned, Padó et al. (2019) have developed relatively simple embedding-based models for claim identification and classification. The experiments presented in this paper are based on their model architecture. As for party manifestos,

NLP methods have been developed for automatic topic analysis (Glavaš et al., 2017). They have been investigated in a comparative fashion due to the textual and conceptual similarity to parliamentary speeches, which creates fertile grounds for domain adaptation (Daumé III, 2007) and cross-topic argument mining (Stab et al., 2018). For example, Abercrombie et al. (2019) apply the annotation scheme of the MARPOR project[3] to a corpus of parliamentary speeches to automatically label policy preferences. To the best of our knowledge, this work is the first study that attempts to establish a direct comparison between manifestos and newspaper reports of the political debates while being grounded in the application of automatic classification methods.

## 3 Methodology

In what follows, we spell out the two steps of our methodology underlying the experiments presented in the subsequent sections.

**Step 1: Automatic claim detection** For our machine learning experiments, we employ the political claim detector from Dayanik and Padó (2020). It models claims detection as a binary classification task at the sentence level where the goal is to decide whether each input sentence contains a claim or not. The architecture is shown in Figure 2. It is based on the BERT architecture (Devlin et al., 2019) which is used to generate sentence representations by computing an embedding for the special `[CLS]` token used to indicate sentence breaks. We use a language specific BERT that is trained on German corpora[4] since it is better at finding subword units for German than the multilingual BERT model (Rönnqvist et al., 2019). A softmax classifier is then placed on top of BERT. It takes the `[CLS]` embedding performs the claim/no-claim classification.

The model is trained using the DEbateNet-mig15 (Lapesa et al., 2020) data set. This data set contains about 2.000 claims from over 450 articles on the domestic discourse of migration in the German quality newspaper *taz* in the year 2015. We use 10 random train, development, and test splits and report the average. Following recommendations made by Devlin et al. (2019), we use the Adam optimizer with learning rates of $5{\cdot}10^{-5}$, $\beta_1 = 0.9$, $\beta_2$

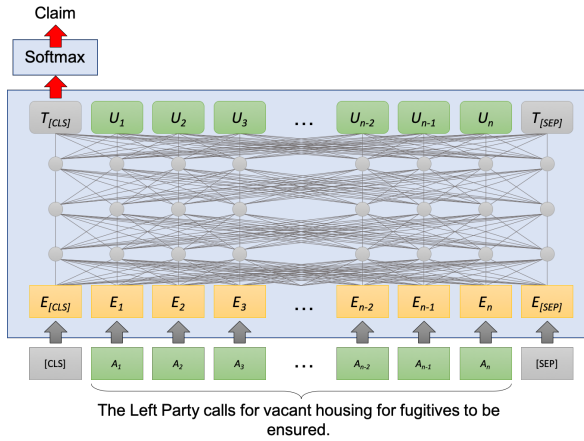---

[3]Manifesto Research on Political Representation, https://manifesto-project.wzb.eu/

[4]https://deepset.ai/german-bert

Figure 2: Model architecture of political claims detector (from Dayanik and Padó, 2020)

| party | # analyzed text spans | # annotated claims | ratio |
|---|---|---|---|
| AfD | 118 | 63 | 0.53 |
| CDU | 114 | 63 | 0.55 |
| Green | 427 | 276 | 0.65 |
| Left | 257 | 159 | 0.62 |
| SPD | 247 | 161 | 0.65 |
| total | 1163 | 722 | 0.62 |

Table 1: Annotation statistics on the Manifesto data set (ratio: claim "density")

= 0.999, a batch size of 32, a gradient clip threshold of 1.0, dropout with p=0.1, and 15 epochs. With this configuration, our model achieves an F1 score of 52.2 on the debate test set (P=40.1, R=74.7). This classifier will be applied and evaluated in Section 4.2.

**Step 2: Semi-Automatic Analysis for Discourse Network Construction.** The NLP approach to the evaluation of a classifier just outlined assumes that the final goal is to detect every single claim correctly. However, in the political science community it has been argued recently that the core of a debate can be perfectly captured even from imperfectly analyzed data (Haunss et al., 2020). An important factor is *redundancy*: the core claims of debates tend to be mentioned multiple times in an article, and thus not every occurrence must be identified. This is particularly true for newspaper reporting, but also holds for party manifestos. This realization motivates our attempt, even in the face of a relatively modest performance of the claim classifier, to construct *discourse networks* of political debates based on newspaper and manifesto texts (cf. Section 1).

That being said, the classifier described above only provides part of the information necessary to create discourse networks. We therefore set up a semi-automatic analysis procedure which we will apply in Section 4.3. First, the researcher runs the automatic claim identifier. They then manually add the polarity (agreement or rejection of claim), identify the actor (in the case of manifestos, this is always the party), filter the false positives from the automatically identified suggestions, and categorize

the remaining claims according to the codebook of Lapesa et al. (2020). Under the assumption that claim identification is the most time-consuming of the steps mentioned above our hypothesis is that this semi-automatic procedure allows us to save a significant amount of time and effort. Indeed, analysis proceeds faster because it is not necessary to read large portions of text (or, potentially, entire articles) which do not contain any claims.

## 4 Experiments

This section introduces the manifesto data (Section 4.1) and presents results for fully automatic model performance in the cross-text type setting (Section 4.2). We then carry out a more substantive comparison in terms of discourse networks and party positions that combines automatic analysis with manual post-processing (Section 4.3).

### 4.1 Manifesto Data Set

The manifesto data set we use encompasses the electoral programs of five German parties from the preceding (2013) and succeeding (2017) election campaigns as collected by the MARPOR project (Volkens et al., 2019).[5] We restricted ourselves to those manifesto sections that dealt with migration-related topics. We built on the MARPOR segmentation of the manifestos into text spans, which often, but not always, correspond to sentences.

We re-used the annotation scheme and pre-defined categories ("codebook") proposed by Lapesa et al. (2020) for the Manifesto data set in order to transfer and compare results from one

---

[5]We considered all parties represented in the German Bundestag in both legislative periods: The Christian Democratic Union (CDU), the Social Democratic Party (SPD), the Green Party (Green), the Left Party (Left). We added the migration-skeptic Alternative for Germany (AfD).

| | $M \cup D$ | $M \cap D$ | $D - M$ | $M - D$ |
|---|---|---|---|---|
| manual | 98 | 71 | 13 | 14 |
| semi-auto. | 97 | 70 | 14 | 13 |

Table 2: Relationships among sets of claim categories in debates (D) and manifestos (M)

text type to the other. The codebook itself consists of eight higher-level categories (e.g. controlling migration, residency, and foreign policy), which in turn are divided into over 100 smaller sub-categories of political claims (e.g. central accommodation, contributions in kind; cf. Figure 1).

The resulting data set contains 722 sentences enclosing at least one claim (on average 144 per party) spread over 1.163 text spans; Table 1 shows details per party. In other words, roughly every other text span contains a migration claim.[6]

**Comparison to debate data.** The empirical overlap between the text types can be described on two levels: On the level of claim categories per genre and on the level of claim categories per genre *and* political party. Table 2 shows the relationships between sets of unique claim categories in the two data sets both for manual and semi-automatic analysis (see Section 4.3). The first row indicates that 98 distinct claim categories exist in total, from which 71 appear in both formats. 13 are unique to the newspaper debate and 14 only appear in electoral programs. The latter appear to be mostly categories that were added during codebook revisions and which may not have been existent for the full period of the annotation of the debate data. Therefore, the difference may be artificially overestimated. Conversely, claims specific to the debate texts deal, e.g., with acute issues related to first admission and accommodation of refugees that were not deemed general enough for inclusion in electoral manifestos.

## 4.2 Step 1: Automatic Claim Identification

To quantify the performance of automatic claim identification, we report precision, recall, F1-score, and accuracy for the model trained on the 2015 debate corpus used to classify each manifesto corpus text span. We break down results by manifesto year (2013 vs. 2017). We expect systematic misclassifications, given that the move from newspaper

---

[6]The manifesto data set is available as a CLARIN resource at the PID http://hdl.handle.net/11022/1007-0000-0007-E7E7-0.

articles to party manifestos involves shifts both conceptually and linguistically.

Overall, the F1-score ranges on a high level from 0.78 in 2017 to 0.86 in 2013 (cf. Table 3). The model also retains a high recall (0.84 in 2013 / 0.79 in 2017) as well as precision (0.88 / 0.77) and accuracy (81.31 / 73.63).

This is bolstered by the second row of Table 2, which shows that all but one claim categories have been identified by the automatic model at least once. Apparently, the model trained on newspaper debate data transfers well to the manifesto corpus and is able to reliably detect political claims in both text types even though the test data is considerably different from the training data. This is positive and encouraging news; we attribute this primarily to the higher density of claims in the manifestos. Additionally, manifestos seem to express their claims in a more concise and unembellished language compared to newspaper articles, sometimes even using enumerations with claims back to back. As a consequence, the performance obtains a better fit than on the original training data set, especially since we narrowed it down to relevant sections of the manifestos beforehand.

Given that recall is generally lower than precision, a remaining question is why, and in what instances, the detection model fails to identify claims correctly. One reason is the difference in coding units (complete sentences in the debate vs. subsentential spans (sub) in the manifestos, see Table 4 for examples). Another is simply the usage of exclamation marks in electoral programs that are uncommon for claims reported in newspapers ('*Stop the harassment of refugees!*' - Manifesto of the Left party, 2013).

Yet, there are further possible causes. For this we turn to the striking difference in performance in both precision and recall between 2013 and 2017: The claim detection model achieves better results for the time period before the training data than after. One possible explanation is that the focus on the content of the 2013 manifestos carries over into the early stages of the 2015 debate and fades away by the end of the year. To rule out this hypothesis, we trained one detector-model on the discourse data from the beginning of the year up until shortly before the peak of the crisis (January to August), and another from there to the end of the year (September to December). Each model still performed better in terms of F1 score on the electoral

|  | 2013 | | | | 2017 | | | |
|  | recall | precision | F1 | accuracy | recall | precision | F1 | accuracy |
|---|---|---|---|---|---|---|---|---|
| AfD | 0.83 | 1.00 | 0.91 | 85.71 | **0.70** | 0.82 | 0.75 | 76.58 |
| CDU | 0.83 | 0.83 | 0.83 | 80.82 | 0.73 | **0.55** | 0.63 | 53.66 |
| Green | 0.80 | 0.90 | 0.85 | 77.97 | 0.86 | 0.80 | 0.83 | 78.32 |
| Left | 0.88 | 0.91 | 0.90 | 85.71 | 0.82 | 0.77 | 0.80 | 74.52 |
| SPD | 0.90 | 0.86 | 0.88 | 83.78 | **0.70** | 0.77 | 0.73 | 67.05 |
| total | 0.84 | 0.88 | 0.86 | 81.31 | 0.79 | 0.77 | 0.78 | 73.63 |

Table 3: Precision, Recall, and F1-Score for claim identification on manifesto data

programs of 2013 than those from 2017, making a seasonal after-effect unlikely.

This begs the question of where the difference in performance between years stems from: Given that the unit of interest is unchanged and it is plausible to assume that the perceived crisis situation of 2015 left a formative imprint in the succeeding manifestos, one might expect better results for 2017.

One explanation applies to the level of machine learning: statistical models tend to decrease in performance with temporal distance from their training data due to the changes in the underlying distribution, a phenomenon known as concept drift (Gama et al., 2014). Arguably, extrapolating into the future should be more difficult than extrapolating into the past, even though we are not aware of specific studies in NLP on this aspect.

Another set of explanations can however be found in the political circumstances and outcomes accompanying the different elections. As a result of the federal elections in 2013, the government coalition changed and the newly founded far-right AfD entered the political landscape in Germany.

Incidentally, the parties most affected by this are the parties in government (SPD and CDU) and the AfD (see bold numbers in Table 3). We propose three plausible explanations for this behaviour:

- The first one is connected to linguistic changes in electoral programs from 2013 to 2017. The SPD's new position in government is reflected in such changes: a) The precondition of re-election is sometimes omitted, making claims sound factual instead of prospective (mood, see Table 4); and b) we observe an increased use of the passive voice (PV). The AfD, too, uses this deviant linguistic style. Since this is very different from the usual language newspapers use to report demands and intentions of political actors, those claims are not reliably recognized by our models. This is due to the fact that political claims analysis, both by definition and concretely in our coding scheme, requires an attributable actor. As a consequence, our models overlook claims, and recall drops.
- The CDU on the other hand keeps reminding the voter of how well they handled the crisis, signaling a claim to our identifier, even though

| Party | Quote | Class | Reason |
|---|---|---|---|
| AfD | All rejected asylum seekers must be returned to their countries of origin. | FN | PV |
| SPD | [...] asylum procedures will continue to be conducted on European soil. | FN | PV |
| CDU | We have effectively reduced the number of those who do not have the right to stay. | FP | CA |
| CDU | We have helped many people in need and offered them hospitality and shelter. | FP | CA |
| SPD | In this [*proposed*] procedure, people make the application before entering Europe. | FN | mood |
| SPD | In this way, the identity is [*will be*] also determined in advance and registration is carried out. | FN | mood |
| AfD | We [...] demand mandatory age tests in cases of doubt, | TP | sub I |
| AfD | [*and*] the exclusion of family reunification. | FN | sub II |
| CDU | Especially in times of uncertainty, we need public institutions [...] | FP | NM |
| SPD | We are therefore committed to a new disarmament initiative. | FP | NM |

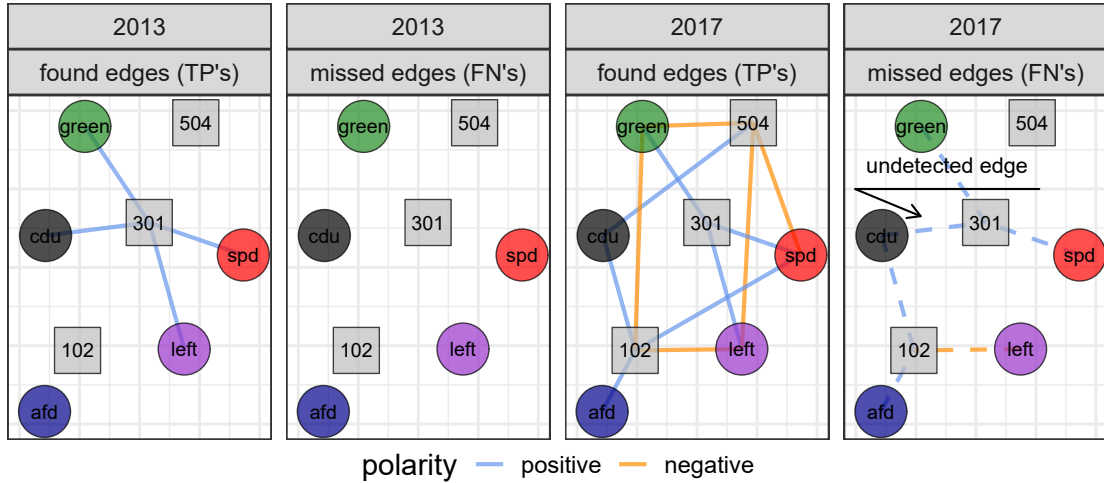Table 4: Examples of misclassifications from 2017

Figure 3: Identifying discourse networks in manifesto data

it is no longer an active demand but rather an already completed action (CA). This increases the number of false positives and results in a lower recall value.

- The most important difference, however, is that across all programs, parties expanded their treatment of migration and the topic started to permeate other subjects: In 2017, migration has become the justification for a lot of claims that were previously unrelated or peripheral (NM, cf. Table 4). Thus, the false positive rate increases and precision declines.

In conclusion, despite considerable textual and conceptual peculiarities, model performance in both years is on a surprisingly high level, modulated by ongoing political shifts.

## 4.3 Step 2: From Claims to Discourse Networks

Given the relatively robust performance of the computational model as well as the high overlap between correctly identified categories, we proceed to apply our model to a more substantive use case, namely the creation of discourse networks (see Figure 3). These networks harbour the potential to – at least partially – explain shifts in party position in their corresponding manifesto with shifts during the debate and vice versa.

Methodologically, discourse network analyses (DNA) aim to reveal discursive patterns of policy debates by focusing on the relations (edges) between actors and claims (vertices, cf. Figure 1) over time (Leifeld, 2016). In other words, the networks tie political to textual entities (concepts) to trace how a political discussion evolves, and how

influence is exerted to shape the direction of the discourse without directly addressing the political counterpart. Therefore, it is also particularly suitable for party manifestos, in which an open exchange of blows is neither possible nor desirable for most parties (Budge et al., 2001).

To carry out a discourse network analysis of the manifesto data, we employ the semi-automatic approach outlined in section 3.

We trace the development of prominent claims (upper limit - 102, integration offers - 301, safe countries of origin - 504) in 2013 and 2017 in Figure 3. Panels 1 and 2 describe the year 2013. Panel 1 shows the edges that are both manually and automatically identified (TPs): apparently only one claim of the three is supported by the four established parties. The empty Panel 2 shows that no such edges have been missed by the model. The picture changes once we move to 2017: Not all edges identified in 2017 (Panel 3) have also been identified each time (dashed lines, Panel 4).

However, as can be seen in the last panel, 5 out of 6 times this bears no consequence for the network topography, because the claim has already been found in another instance. Only in the case of the CDU's support of integration offers, the model failed to detect a claim and thus the network misses an edge. Auspiciously, this is rather the exception than the rule: For all claims in the entire network, only 7 percent of edges are missed by the semi-automatic approach (22 cases out of 333). We thus conclude that the claim identifier trained on newspaper data is a surprisingly good fit for party manifestos, due to the informational redundancy.
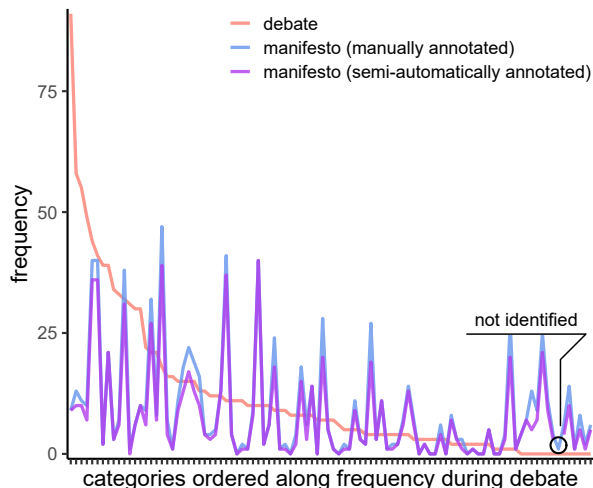
30

Figure 4: Claim category distribution in debates and manifestos



Figure 5: Scatterplot of positions (with jitter for clarity)

## 5 Issue salience and party position

In light of the promising performance, we now turn to the two remaining questions of substantive overlap between newspaper debate and party manifestos from the literature (cf. Section 2).

**Salience.** The first one is: Can we in fact observe a different focus in terms of salience between formats? Figure 4 depicts the frequency distributions of (manually annotated) claim categories across public newspaper debate (red) and manifestos (blue), sorted by debate frequency. We observe considerable differences: popular categories during the 2015 debate are by no means also predominant in the party manifestos of 2013 and 2017. The correlation between genres amounts to only $r = 0.01$ ($p = 0.91$; for 2013 and 2017 alone $r = 0.28$ and $r = 0.1$, respectively), once one takes party specific salience into account. These results align well with findings from the literature on party salience. In contrast to the debate–manifesto split, the manually annotated distribution in the manifestos (blue) correlates extremely highly with the semi-automatic distribution (purple). ($r = 0.99$, $p < 0.01$), carrying over to individual parties as well, ranging from $r = 0.91$ to $r = 0.98$. This underlines again the usefulness of automatic analysis.

**Positions.** A final question is whether the positions of parties regarding certain categories are similar across formats and whether a semi-automatic approach captures complementary tendencies. Po-
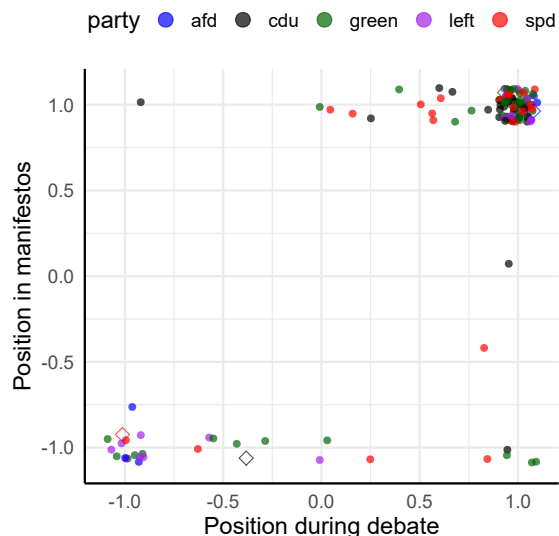
sitions ($P$) can be calculated as the difference between all positive and negative referrals to a certain claim $C$ divided by their sum (Kim and Fording, 2003, p. 97-98):

$$P = \frac{(\sum C_{pos} - \sum C_{neg})}{(\sum C_{pos} + \sum C_{neg})} \quad (1)$$

Figure 5 shows the positioning of parties on different claims in the 2015 debates (x-axis) as well as the union of their 2013 and 2017 manifestos (y-axis), ranging from -1 (total opposition) to 1 (total support). This translates into a scale with the endpoints acceptance or rejection regarding a certain political demand with more moderate positions in between. The four uncoloured squares mark claims which were not automatically identified for this particular party in their manifesto. Most observations are aligned perfectly across genres and located in two clusters (bottom left and top right corner). This indicates consistent party positions between text types (and also years) with occasional moderate changes and very few volte-faces.

Again, this confirms findings from the literature: The positions derived from manifestos are a strong indicator of positioning during discourse and vice versa ($r = 0.79$, $p < 0.01$; for 2013 and 2017 alone $r = 0.69$ and $r = 0.83$, respectively). Unsurprisingly, we observe more variance in the position of parties during the debate (x-axis, Figure 5), since it displays the aggregated and at times contradictory position of individual party members

instead of an unified party.

It is also noteworthy that 62% of party specific claims appear in one text type but not the other. This fact diminishes the party-specific overlap between formats quite a lot and also impacts the meaningfulness of the correlation. Interestingly enough, the correlation between manually and semi-automatically estimated positions (excluding the missing claims) appears unaffected by this bias and amounts to $r = 0.99$ ($p < 0.01$). Analogous to the application scenario in 4.3, this can to a great extend be explained by textual redundancy: Assuming that parties emphasize claims multiple times and streamline their texts to avoid contradictions, we only need to find a single repetition of a specific claim to find the correct representation of that position. That makes a semi-automatic approach to positions much less demanding than to salience, because in the latter case every instance matters. In summary, while parties did not put the same emphasis on categories across text types, their standpoints appear to be consistent over the short time window under investigation.

## 6   Conclusion

In this paper, we have explored whether computational models for claim detection trained on newspaper data could be applied to party manifestos as well. We found both high numeric performance, with F1 scores in the range of 0.80, and substantive overlap between manual and semi-automatically annotated sections in electoral programs.

A first surprise was that the claim detection model achieved better results for the 2013 than for the 2017 manifesto. We have offered a machine learning-based explanation, and a set of three political ones: The first two were connected to deviant linguistic styles, and the third to the increasingly ubiquitous nature of the migration topic after the so-called 'refugee crisis' of 2015. These hypotheses can clearly benefit from a future more systematic re-examination on a larger scale.

Assessing the perspectives for building discourse networks from the party manifestos, we were able to semi-automatically identify 93% of the edges by taking advantage of textual redundancy. With this in mind, we were able to confirm existing findings from the political science literature on party position and issue salience. While the positions of different parties on distinct claims are mostly congruent between formats, salience turned out to be largely uncorrelated.

Clearly, our current approach has further limitations. First, we only focus on one of many potential policy issues (migration) within a very short time-frame. This begs the question of generalizability, which we aim to address in three ways in future research: Concretely, ongoing work extends our analysis to different years (2005 and 2010), to the issue of pensions, and to another newspaper to validate our findings as well as to account for issue-specific characteristics (such as the degree of polarization or the granularity of proposed policy instruments).

Second, even though we already reduced it considerably, we still rely on human intervention for the semi-automatic annotation. Yet, from a mixed-methods perspective, this approach highlights the mutual benefit of interleaving political science and NLP. On the one hand the cross-validation of discourse networks with networks based on electoral programs could open up new venues for DNA and, conversely, explore the role public discourse plays in the formation of these programs. On the other hand, a still open question is whether training and test data can be reversed. To answer this question additional labeled material is needed. Positive results might prove as a welcome shortcut to the creation of comprehensive codebooks for political claims analysis given the much higher ratio of claims to text in manifestos. Both approaches would require computational assistance and thus enable further integration of NLP and CSS.

## Acknowledgments

## References

Gavin Abercrombie, Federico Nanni, Riza Batista-Navarro, and Simone Paolo Ponzetto. 2019. Policy Preference Detection in Parliamentary Debate Motions. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 249–259, Hong Kong, China. Association for Computational Linguistics.

James Adams. 2001. *Party Competition and Respon-*

*sible Party Government*. University of Michigan Press.

American Political Science Association. 1950. *Toward a More Responsible Two-Party System. A Report of the Committee on Political Parties*. Menasha, Wis.

Judith L. Bara. 2006. The 2005 Manifestos: A Sense of Déjà Vu? *Journal of Elections, Public Opinion and Parties*, 16(3):265–281.

Kenneth Benoit and Michael Laver. 2007. Estimating party policy positions: Comparing expert surveys and hand-coded content analysis. *Electoral Studies*, 26(1):90 – 107.

Ian Budge. 1987. The internal analysis of election programmes. In David Robertson, Derek Hearl, and Ian Budge, editors, *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*, pages 15–38. Cambridge University Press, Cambridge.

Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Tanenbaum. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments, 1945-1998*. Oxford University Press.

Elena Cabrio and Serena Villata. 2018. Five Years of Argument Mining: A Data-driven Analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 5427–5433, Stockholm, Sweden. International Joint Conferences on Artificial Intelligence Organization.

Hal Daumé III. 2007. Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, Prague, Czech Republic. Association for Computational Linguistics.

Erenay Dayanik and Sebastian Padó. 2020. Masking actor information leads to fairer political claims detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4385–4391, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. 2014. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4).

Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. 2017. Cross-lingual classification of topics in political texts. In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 42–46, Vancouver, Canada. Association for Computational Linguistics.

Christoffer Green-Pedersen and Peter B. Mortensen. 2015. Avoidance and Engagement: Issue Competition in Multiparty Systems. *Political Studies*, 63(4):747–764.

Martin Haselmayer, Thomas M Meyer, and Markus Wagner. 2019. Fighting for attention: Media coverage of negative campaign messages. *Party Politics*, 25(3):412–423.

Sebastian Haunss, Nico Blokker, Andre Blessing, Erenay Dayanik, Gabriella Lapesa, Jonas Kuhn, and Sebastian Padó. 2020. Integrating manual and automatic annotation for the creation of discourse network data sets. *Politics and Governance*, 8(2):326–339.

Marc Helbling and Anke Tresch. 2011. Measuring party positions and issue salience from media coverage: Discussing and cross-validating new indicators. *Electoral Studies*, 30(1):174–183.

Heemin Kim and Richard C. Fording. 2003. Voter ideology in Western democracies: An update. *European Journal of Political Research*, 42(1):95–105.

Ruud Koopmans and Paul Statham. 1999. Political Claims Analysis: Integrating Protest Event and Political Discourse Approaches. *Mobilization: An International Quarterly*, 4(2):203–221.

Gabriella Lapesa, Andre Blessing, Nico Blokker, Erenay Dayanik, Sebastian Haunss, Jonas Kuhn, and Sebastian Padó. 2020. DEbateNet-mig15:tracing the 2015 immigration debate in Germany over time. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 919–927, Marseille, France. European Language Resources Association.

Philip Leifeld. 2009. Die Untersuchung von Diskursnetzwerken mit dem Discourse Network Analyzer (DNA). In Volker Schneider, Frank Janning, Philip Leifeld, and Thomas Malang, editors, *Politiknetzwerke. Modelle, Anwendungen Und Visualisierungen*, pages 391–404. VS Verlag für Sozialwissenschaften, Opladen.

Philip Leifeld. 2016. *Policy Debates as Dynamic Networks: German Pension Politics and Privatization Discourse*. Campus Verlag, Frankfurt/New York.

Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In *Proceedings of the 24th International Conference on Artificial Intelligence*, page 185–191. AAAI Press.

Marco Lippi and Paolo Torroni. 2016. Argument mining from speech: Detecting claims in political debates. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, page 2979–2985. AAAI Press.

Peter Mair. 2009. Representative versus responsible government. Working Paper 09/8, MPIfG working paper.

Gary Marks, Liesbet Hooghe, Marco R. Steenbergen, and Ryan Bakker. 2007. Crossvalidating data on party positioning on European integration. *Electoral Studies*, 26(1):23–38.

Nicolas Merz. 2018. *The Manifesto-Media Link: How Mass Media Mediate Manifesto Messages*. Ph.D. thesis, Humboldt-Universität zu Berlin, Kultur-, Sozial- und Bildungswissenschaftliche Fakultät.

Catherine E. Netjes and Harmen A. Binnema. 2007. The salience of the european integration issue: Three data sources compared. *Electoral Studies*, 26(1):39–49.

Sebastian Padó, Andre Blessing, Nico Blokker, Erenay Dayanik, Sebastian Haunss, and Jonas Kuhn. 2019. Who Sides with Whom? Towards Computational Construction of Discourse Networks for Political Debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2841–2847, Florence, Italy. Association for Computational Linguistics.

Colin Rallings. 1987. The influence of election programmes: Britain and Canada 1945–1979. In David Robertson, Derek Hearl, and Ian Budge, editors, *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*, pages 1–14. Cambridge University Press, Cambridge.

Leonard Ray. 2007. Validity of measured party positions on european integration: Assumptions, approaches, and a comparison of alternative measures. *Electoral Studies*, 26(1):11 – 22.

David Bruce Robertson. 1976. *A Theory of Party Competition*. J. Wiley.

Samuel Rönnqvist, Jenna Kanerva, Tapio Salakoski, and Filip Ginter. 2019. Is multilingual BERT fluent in language generation? In *Proceedings of the First NLPL Workshop on Deep Learning for Natural Language Processing*, pages 29–36, Turku, Finland. Linköping University Electronic Press.

Tobias Schwarzbözl, Matthias Fatke, and Swen Hutter. 2019. How party-issue linkages vary between election manifestos and media debates. *West European Politics*, pages 1–24.

Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. 2018. Cross-topic Argument Mining from Heterogeneous Sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium. Association for Computational Linguistics.

Jacques J. A. Thomassen. 1994. Empirical Research into Political Representation: Failing Democracy or Failing Models? In M. Kent Jennings and T.E. Mann, editors, *Elections at Home and Abroad, Essays in Honor of Warren Miller*, pages 237–265. Michigan University Press.

Andrea Volkens and Judith Bara. 2013. Presidential Versus Parliamentary Representation: Extending the Manifesto Estimates to Latin America. In Andrea Volkens, Judith Bara, Ian Budge, Michael D. McDonald, Robin Best, and Simon Franzmann, editors, *Mapping Policy Preferences From Texts: Statistical Solutions for Manifesto Analysts*, chapter Mapping Policy Preferences From Texts, pages 277–298. Oxford University Press.

Andrea Volkens, Werner Krause, Pola Lehmann, Theres Matthieß, Nicolas Merz, Sven Regel, and Bernhard Weßels. 2019. The Manifesto data collection, version 2019b.