Come Hither or Go Away?
Recognising Pre-Electoral Coalition Signals in the News

Ines Rehbein
University of Mannheim

Simone Paolo Ponzetto
University of Mannheim

Heiner Stuckenschmidt
University of Mannheim

{ines,simone,heiner}@informatik.uni-mannheim.de

Anna Adendorf
University of Mannheim

Oke Bahnsen
University of Mannheim

Lukas Stoetzer
Humboldt University Berlin

{anna.adendorf,oke.bahnsen}@mzes.uni-mannheim.de

lukas.stoetzer@hu-berlin.de

Abstract

In this paper, we introduce the task of political coalition signal prediction from text, that is, the task of recognizing from the news coverage leading up to an election the (un)willingness of political parties to form a government coalition. We decompose our problem into two related, but distinct tasks: (i) predicting whether a reported statement from a politician or a journalist refers to a potential coalition and (ii) predicting the polarity of the signal – namely, whether the speaker is in favour of or against the coalition. For this, we explore the benefits of multi-task learning and investigate which setup and task formulation is best suited for each sub-task. We evaluate our approach, based on hand-coded newspaper articles, covering elections in three countries (Ireland, Germany, Austria) and two languages (English, German). Our results show that the multi-task learning approach can further improve results over a strong monolingual transfer learning baseline.

1 Introduction

In the last decade the capabilities of computational methods for text analysis have drastically evolved and the usage of Natural Language Processing (NLP) methods have increasingly expanded in scope, enabling a wide range of applications for the automatised analysis of political texts (Glavaš et al., 2019). Much work in NLP in the past has focused on making sense of how legislative processes are codified in text – e.g., by analysing debate motions (Abercrombie et al., 2019; Abercrombie and Batista-Navarro, 2020; Sawhney et al., 2020, *inter alia*). However, when analysing electoral dynamics it seems necessary to complement the application of NLP methods on domain-specific texts such as electoral manifestos with a broader analysis of signals from news texts (Blokker et al., 2020).

Specifically in the case of government formation in multi-party systems – that is, in countries where the government is typically formed by multiple parties – party members and candidates often communicate in interviews and rallies during election campaigns which coalition their party might or might not be willing to enter after the election. These public statements about prospects of future governments are often picked up by media outlets. Being able to automatically identify coalition signals in news coverage has accordingly the potential to enable large-scale studies of voting behaviour using 'text-as-data', with the promise of improving our understanding of voters' behaviour and of strategic voting in multi-party democratic systems (Bahnsen et al., 2020; Gschwend et al., 2017; Stoetzer and Orlowski, 2020). Very recent work in political science made first steps in this direction by investigating dictionary-based approaches to automatically detect coalition signals in newspaper text (Bowler et al., 2021). However, much work still remains to be done by developing data-driven methods that take advantage of advances in our field from the past years such like pre-trained language models and joint learning, and providing a thorough evaluation of the accuracy of such automated approaches. This is especially important because coalition signals can be expressed in text by explicit or implicit statements and make heavy use of irony and figurative language, thus raising many challenges that cannot be solved with simple dictionary-based methods and require Natural Language Understanding (NLU) capabilities.

In this paper, we present the task of coalition signal detection as a new, challenging NLU task. We propose to approach this task as a machine learning problem and decompose it into two subtasks, (i) the identification of text sequences that contain a signal and (ii) the prediction of the polarity of each potential signal, that is, whether the speaker promotes or opposes the signalled coalition – cf. examples 1.1 and 1.2 for sentences containing a coalition signal and negative polarity (i.e., unwillingness to join a

coalition), respectively.

Example 1.1. The Green Party yesterday signalled that it would be interested in participating in a coalition government led by Fine Gael.

Example 1.2. Ms O'Donnell said she could not see the PDs negotiating with the Labour Party after the election.

We benchmark performance on this task by introducing a new multilingual dataset – using pre-election newswire coverage from multiple elections in three countries (Ireland, Germany and Austria) – and quantifying the performance of monolingual pre-trained models to provide a simple, yet state-of-the-art strong baseline that is shown to substantially outperform previous approaches from the text-as-data community in political science. Moreover, we investigate the benefits of auxiliary tasks and a cross-lingual approach in a multi-task learning setting, so as to improve performance over the simpler, transfer-learning-based baseline.

2 Related Work

In recent years the usage of computational methods for the analysis of 'text-as-data' has drastically expanded in scope (Grimmer and Stewart, 2013; Glavaš et al., 2019). Political scientists have used NLP methods for a wide range of tasks, including, for instance, inferring policy positions (Lowe et al., 2011) and detecting topics (Grimmer, 2010). At the same time, NLP researchers have addressed closely related tasks such as election prediction (O'Connor et al., 2010), agreement measurement (Menini et al., 2017) and claim detection (Dayanik and Padó, 2020) to name a few.

When focusing on inter-party communication, researchers have used information from the news (De Nooy and Kleinnijenhuis, 2013; Elmelund-Præstekær, 2008; Lau and Pomper, 2002) or have investigated different kinds of campaign ads (Damore, 2002; Elmelund-Præstekær, 2011; Walter and Van der Brug, 2013), press releases (Dolezal et al., 2016, 2017; Haselmayer and Jenny, 2018), election manifestos (Curini, 2011; Dolezal et al., 2018), or social media data (Auter and Fine, 2016; Gervais et al., 2020; Gross and Johnson, 2016).

This type of analysis typically requires manual coding of theoretical concepts and has been used to study the role of coalition signals and negative campaigning in the context of electoral campaigns (Best, 2015; Curini and Martelli, 2010; Druckman et al., 2009; Elmelund-Præstekær, 2010; Evans et al., 2014; Walter and Van der Eijk, 2019). Manual coding, however, is a labour and time consuming process which makes it hard to apply to larger amounts of data or to longitudinal studies where new data will be collected over a long time period, thus requiring new manual effort for the same coding steps. It is therefore crucial to automate this process in order to enable large-scale studies that might increase our understanding of the role of coalition signals in electoral campaigns.

One new promising approach towards that end was recently presented by Bowler et al. (2021), who developed a dictionary-based method for detecting inter-party communication in the form of pre-electoral coalition signals. Whilst their approach so far only identifies positive signals between parties, i.e., parties indicating that they are willing to form a government coalition together, it is a promising starting point to automate the process of studying inter-party communication. We describe their approach in more detail in §4.3 and then present our own approach to the identification of inter-party communication in the form of coalition signals during electoral campaigns (§4.2).

3 Task definition and annotation

Data Creation and Annotation. Our data includes election coverage from Irish, German and Austrian daily newspapers (Ireland: Irish Times, Irish Independent; Germany: Süddeutsche Zeitung (SZ), Frankfurter Allgemeine Zeitung (FAZ); Austria: Der Standard, Die Presse). SZ and FAZ are considered the two most high-quality/high-circulation German daily newspapers (see, e.g., Weischenberg et al. (2006)). The Irish/Austrian newspapers were selected according to similar considerations. All articles were published in a period up to four weeks (28 days) before the election, not including the actual election day. From that time period, we sampled relevant articles based on the following search query:

(1) ```
     election AND (*coalition* OR
     pact OR *collaboration* OR
     alliance)[1]
     ```

The keyword-based sampling was necessary so that the coders did not spend too much time on articles that did not contain any signals. The keywords were optimised for recall by domain experts who,

---

[1] For German we used the following query terms: Wahl (election), Koalition (coalition), Bündnis (pact), Zusammenarbeit (collaboration), Allianz (alliance).

during keyword selection, also checked articles without any hits for false negatives, to avoid missing relevant articles.

The resulting German dataset is the largest with 15,735 sentences of newswire text. The Irish and Austrian data are smaller with 8,493 and 2,880 sentences. The extracted articles have been filtered manually for false positives, i.e., articles that include the keywords but do not touch on the issue of upcoming elections have been removed. As a result, each article in the corpus includes at least one coalition signal.[2]

**Coding coalition signals.** In order to code coalition signal in newspaper texts, the annotators[3] first read the whole article and highlighted all signals and speculations in each article (cf. definitions below). After this initial step, the annotators proceeded with the actual coding: all highlighted signals and speculations in the article were coded in chronological order and inserted in the coding template. For each signal or speculation, the sender(s) and addressee(s) are identified, so that each can be attributed to one or more specific parties. In cases where this information is missing from the text, the instance is coded as *unspecified*. Crucially, our annotation scheme distinguishes coalition *signals* and *speculation*.

**Definition 3.1** (SIGNAL). A coalition *signal* is a statement made by a potential coalition partner about their preferences towards a possible coalition in which the party itself would be a member.

A coalition signal consists of three components:

(1) a SENDER (the potential coalition partner)

(2) an ADDRESSEE (the potential coalition)

(3) a *positive, neutral* or *negative* statement about the potential coalition.

For being identified as a coalition signal, it is required that the sender directly speaks about their preferences concerning a potential coalition. This includes citation of (direct and indirect) quotes by a journalist. Sometimes the addressee is *underspecified* (e.g., when a party rules out a coalition with "any left party").

We present a few examples to illustrate instances of coalition signals. Example 3.1 includes a *positive* signal from Mary Harney (Independent) (SENDER) to Fianna Fáil (ADDRESSEE). A *negative* signal is shown in 3.2. Here, the SENDER are the Social Democrats and the ADDRESSEES are Fine Gael and the Labour Party. Example 3.3 contains a neutral signal from Enda Kenny (Fine Gael) to the Labour Party.

**Example 3.1.** Minister for Health Mary Harney said she had no problem about remaining in the Government with Fianna Fáil at present.

**Example 3.2.** But meanwhile, Mr Donnelly last night played down the possibility that the Social Democrats would be willing to go into a coalition with Fine Gael and the Labour Party.

**Example 3.3.** Asked if the row between Fine Gael and Labour had now become so acrimonious as to threaten the possibility of the two parties working together in a coalition arrangement, Mr Kenny did not give a direct reply.

**Coding coalition speculation.** Speculations are typically stated by journalists, experts and political actors. Like coalition signals, each speculation has a SENDER, an ADDRESSEE and a statement regarding the anticipated coalition, with positive, negative or neutral polarity.

**Definition 3.2** (SPECULATION). A coalition *speculation* is a statement about someone else's (i.e., not the speaker's) preferences concerning the formation of a coalition after the election.

Example 3.4 speculates about a coalition of Fine Gael and Labour, with positive polarity. As the roles of SENDER and ADDRESSEE are not made explicit, we code two separate instances of speculation (one with Fine Gael as SENDER and Labour as ADDRESSEE, and one where the roles are reversed). Example 3.5 includes two negative speculations from Fine Gael (SENDER), the first one directed to Fianna Fáil and the second one to Sinn Fáin.[4] Finally, example 3.6 illustrates a neutral coalition statement from the Greens (SENDER) to Fianna Fáil (ADDRESSEE).

**Example 3.4.** The likelihood of Fine Gael and the Labour Party beginning coalition talks in just over

---

[4]This poses an additional challenge for an automated system that needs to distinguish signals/speculation with more than two addressees from multiple signals/speculation with only one addressee (i.e., FG+FF+SF vs. FG+FF, FG+SF).

two weeks' time is as close to a sure thing as exists in politics.

**Example 3.5.** The only parties Fine Gael had ruled out as coalition partners were Fianna Fáil and Sinn Fáin.

**Example 3.6.** The party, which is fielding seven candidates including Cllr Eamon Ryan in Dublin South and Cllr Ciarán Cuffe in Dún Laoghaire, did not rule out coalition with Fianna Fáil after the election.

Arguably, our examples show how challenging our task is, since it requires information on speaker attribution, coreference resolution and polarity shifting (e.g., example 3.6 would shift its polarity if the statement would be *did rule out* instead of *did not*). Moreover, the task requires a considerable portion of world knowledge (e.g., for example 3.6 it is important to know that Eamon Ryan and Ciarán Cuffe are both members of the Green Party; for the German and Austrian data one needs to know that expressions like "Ampel" (traffic light) and "Wachteleier-Koalition" (quail egg coalition) refer to specific coalition constellations.[5]

**Intercoder reliability** Tests of intercoder reliability concerning the annotations for Germany [Austria, Ireland] yielded a Krippendorff's alpha of 0.595 [0.903, 0.659] for the proportion of signals (among signals and speculations) per article. It is not quite clear why the agreement for the Austrian data is higher than for the other two countries. We can exclude training effects (each country has been coded by a different set of annotators). Our intuition is that this at least partly reflects newspaper style, with less figurative/implicit signals in the Austrian newspaper articles.

**Signals versus speculation** The coding of signals and speculation in our data enables a comparison of the two types of elite communication. For Germany and Ireland, signals were more frequent than speculation for most C-types, with some exceptions. For new or improbable C-types, we observe more speculation than signals (e.g., for the German C-type "Greens + Left + SPD", we count six times more speculations than signals. Taking a closer look at the data and also considering the polarity of the signals, we notice that most of the signals are negative while the vast majority of the

---

[5]*Ampel* describes a coalition between the German SPD, FDP and the Green Party; *Wachteleier-Koalition* refers to a coalition of SPÖ and FPÖ.

speculations are positive. Similar observations can be made for the Irish data where the ratio of signals/speculations for the most frequent C-type (Fianna Fáil + Fine Gael) is 0.6/0.4. Here, too, most of the signals have a negative polarity while the majority of the speculations are positive. These observations suggest that coalition signals and speculation are used differently and that it might be worthwhile to take a closer look at the differences between the two.

## 4 A transfer learning approach to coalition signal detection

We propose an approach to coalition signal detection on the basis of transfer learning, and decompose our problem into two separate classification tasks. The first task consists of identifying whether or not a certain text sequence – more specifically, a sentence – includes a coalition signal. For signal-bearing sentences, we additionally predict the *type* of the coalition (e.g., LABOUR+SOCIAL DEMOCRATS). For this, we model both steps in a standard text classification framework by assigning non-signals the label NONE. The second task then determines the polarity of the detected signals as either *negative, neutral* or *positive*.

### 4.1 Experimental setup

Table 1 shows the distribution of unique coalition types (C-types) in each dataset. In total, we have 100 distinct coalition types coded in the data.

In order to predict the final labels for our data, we collect the sentence-level annotations from both classifiers and merge them into a new, atomic label (e.g., the C-type: FINE GAEL+GREENS+LABOUR and polarity prediction: *positive* would be merged into the atomic label FINE GAEL+GREENS+LABOUR:*positive*). As the number of unique coalition types in the data is already quite high (see Table 1) and most labels are sparse, we focus on the most frequent types and only try to predict those. For that, we determined a different minimum occurrence frequency $N$ for each country, depending on data size: German: N=10, Ireland: N=12, Austria: N=6. As a result, we obtain 14 different coalition types for German elections, 12 distinct types for Irish and 8 types for Austrian elections (Appendix, Table 7) that we use to train our models.

To ensure fair benchmarking of our models with respect to the full labeled data, we include for test-

| | Signal | Spec. | Sent. | C-types |
|---|---|---|---|---|
| **DE** | 443 | 354 | 15,735 | 26 |
| **IE** | 297 | 225 | 8,493 | 50 |
| **AU** | 47 | 57 | 2,880 | 24 |
| **Total** | 757 | 636 | 27,108 | 100 |

Table 1: Distribution of coalition signals and speculation in the three datasets (Sent: no. of sentences, C-types: unique coalition types in the gold standard).

| | C-type | | | C-type+pol | |
|---|---|---|---|---|---|
| | orig | mod | train | pred | gold |
| Germany | 39 | 26 | 14 | 42 | 58 |
| Ireland | 80 | 50 | 12 | 36 | 28 |
| Austria | 32 | 24 | 8 | 27 | 16 |

Table 2: No. of unique **C-type labels** in the original data (orig), in the modified data (mod), in the training data (train) and number of unique **C-type+pol labels** in the system predictions (pred) and test set (gold).

ing, however, all coalition signals, i.e., *all* coalition types that have been coded in the data, including those not appearing in the training set.

We compare our approach to the dictionary-based approach of Bowler et al. (2021) where the authors use a dictionary with party names and terms indicating cooperation or coalition to identify sentences in news coverage of German elections that might include coalition signals.[6] If a sentence includes a term signaling cooperation *and* a reference to a coalition, it is classified as a signal. The criteria for coalition references are as follows: The sentence must include at least two mentions of different parties *or* a coalition term that can be resolved to the parties participating in the coalition (e.g.: "große Koalition" (grand coalition) $\rightarrow$ CDU/CSU + SPD).

It is important to note that Bowler et al. (2021) do not distinguish *signals* from *speculation* and also do not distinguish between sender and addressee but use the same label for signals from party A to B as for signals from party B to A. We therefore modify our data to make our results comparable. We map SENDER and ADDRESSEE information to undirected coalition terms (e.g., both SENDER: Labour, ADDRESSEE: Social Democrats and SENDER: Social Democrats, ADDRESSEE: Labour are mapped onto LABOUR+SOCIAL DEMOCRATS). In addition, we merge signals and speculation into one unified class. In the remainder of the paper, we refer to both, signals and speculation, as *signals*.

As mentioned before, Bowler et al. (2021) do not model the signals' polarity. We thus compare results for C-type prediction only, in addition to predicting the full label, C-type+polarity, where we consider all signals identified by the dictionary-based approach as positive signals.

Table 2 summarises the properties of our data

and modifications described above (i.e., removing the distinction between sender and addressee and merging C-type labels with signal polarity). The first column (orig) shows the number of unique C-type labels in the original dataset (encoding the direction of the signal). Column 2 (mod) shows the number of distinct labels after removing the information on sender and recipient (i.e., the direction). As this number is still quite high and many labels are sparse, we focus on the most frequent C-types and only try to predict those. Specifically, we train our models to predict 14 different C-types for German elections, 12 distinct types for Irish and 8 types for Austrian elections (column 3; train).

Column 4 lists the number of merged labels (C-type+polarity; pred) that our models are able to predict while the final column shows the unique number of merged final labels in the ground truth data. The number of predicted label types can be higher than the actual number of labels in the gold annotations as not every possible combination of C-type and polarity has been seen in the data.

## 4.2 Our approach

We propose a transfer learning approach for the task of coalition signal detection, based on pretrained transformer-based language models (Vaswani et al., 2017). For German, we evaluate monolingual BERT (Devlin et al., 2019)[8] and finetune two separate transformer models for each dataset in order to predict (i) the coalition type and (ii) the polarity of the signals.

**Training details and hyperparameters** We use the same model architecture for both subtasks and initialise parameters with the pretrained bert-base-german-cased model, using the same seed in all experiments to control for random effects caused by parameters that remain uninitialised for some

---

[6]The authors provide an R script[7] for replication, based on the QUANTEDA library (Quantitative Analysis of Textual Data) (Benoit et al., 2018).

[8]We use the pretrained bert-base-german-cased model.

of the models. Our implementation is based on the Huggingface transformers library (Wolf et al., 2020) and PyTorch (Paszke et al., 2017), using the AutoModelForSequenceClassification model that can be initialised with different pretrained transformers and has a linear layer on top of the pooled output. As we are facing a sparse and imbalanced data problem where the majority of the sentences in our data do not contain any signal (Table 1), we apply downsampling and report results in a 5-fold cross-validation setup. All our models were trained with a maximum sequence length of 128 tokens per input sentence and a batch size of 8. Initial learning rate was set to 1e-5 and weight decay to 0.01. We did not freeze any of the layers but fine-tuned the whole model in all experiments. All our models are trained on a single GPU with 11GB RAM (NVIDIA RTX 2080Ti).

### 4.3 Dictionary-based baseline

**Evaluation**  One problem for evaluation is the fact that coalition signals can consist of text segments with multiple sentences, while the system predictions (for Bowler et al. (2021) as well as in our task formulation) are on the sentence level. Example 4.1 (translated from German) illustrates how the information needed for disambiguation can be spread over multiple sentences, where the first sentence suggests that Bütikofer, a member of the Green Party, sends a signal to the SPD and FDP (Ampel coalition). The second sentence, however, clarifies that the addressee is the SPD only.

**Example 4.1.** Bütikofer used the image: "We are only available for one traffic light. And that is the pedestrian traffic light, which only has red and green."

A sentence-level evaluation, thus, might be too fine-grained, since it requires the system to identify all sentences that belong to one signal. Accordingly, we evaluate at the signal level by creating a mapping between sentences and coalition signals that allows us to map all predictions for individual sentences back to their corresponding signals. As a result, we can have multiple predictions for each signal. We can then count the number of true positives (correctly identified signal), false positives (system predicts an incorrect signal) and false negatives (system fails to identify the signal) and report standard precision, recall and f-measure:

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad (1)$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \quad (2)$$

### 4.4 Monolingual signal prediction (German)

We evaluate our approach on German (Table 3) and compare our results to previous work (Bowler et al., 2021). Our transformer-based approach significantly outperforms the baseline in each setup.

The first setup (A) evaluates on the level of individual sentences whether a coalition signal has been detected (but not whether the type and polarity of the signal have been predicted correctly). While the baseline has a higher precision than the neural classifier, this comes at the cost of recall (28.9% versus 60.3%). Setting (B) compares the predicted coalition types on the signal level but ignores the polarity of the signals. This is arguably a fairer comparison than (C) where we also consider the polarity of the signal, as the approach of Bowler et al. (2021) does not explicitly model the polarity of the signals. For (C), the transfer learning approach is able to correctly predict coalition types+polarity for more than twice the number of signals than the baseline while keeping precision in the same range (46.3% versus 46.4%). This is a respectable result for such a challenging task where we have a high number of often sparse labels and where the model can not rely on lexical cues only.

## 5 Multilingual coalition signal prediction

We next focus on the prediction of coalition signals in a multilingual setup and across countries. As shown in Table 1, the size of the Irish and Austrian data is much smaller and does not provide a sufficient number of instances for training accurate monolingual predictors. Our main research goal is thus to find appropriate settings for improving the accuracy for coalition signal detection in Irish and Austrian election coverage by making use of the larger German hand-coded data.

Given the relatedness of our tasks, we propose a multitask learning (Ruder, 2017, MTL) approach for our problem. From a conceptual point of view it is important, however, to note also the differences between our two tasks. While they can be both approached as text classification problems, polarity prediction has a small number of labels only (*positive, neutral, negative*) that are the same across countries and languages. Coalition type prediction has instead a much larger number of (often sparse)

|  | (A) signal (yes/no) | | | (B) C-type | | | (C) C-type + polarity | | |
|---|---|---|---|---|---|---|---|---|---|
|  | prec | rec | f1 | prec | rec | f1 | prec | rec | f1 |
| Baseline | 36.04 | 28.89 | 32.07 | 81.01 | 37.66 | 51.41 | 46.29 | 21.02 | 28.92 |
| BERT | 31.38 | 60.27 | **41.28** | 68.52 | 61.24 | **64.68** | 46.38 | 43.13 | **44.69** |

Table 3: Results for the prediction of coalition signals in German newspaper text: (A) identification of coalition signals on the sentence level (binary: yes/no); (B) the prediction of the correct coalition type (C-type) on the signal level; and (C) the prediction of the correct C-type *and* polarity on the signal level.

labels that are country-specific and thus pose a challenge for crosslingual transfer.

## 5.1 Experimental setup

Our MTL setup uses hard parameter sharing (Caruana, 1993) which has shown to work well for related tasks. All models in our setup share the same encoder, initialised with pretrained multilingual MBERT or R-XLM language models, and update the same encoder weights during training.[9] This approach enables us to use the task-specific implementations of the Huggingface library for the different auxiliary tasks.

We test different MTL setups where we model the prediction of a) coalition types and b) signal polarity for different countries as different tasks. We expect that this will allow the model to jointly learn shared feature representations across countries, while the task (i.e., country)-specific classifier will be fine-tuned for each task separately. This seems crucial especially for the prediction of coalition types where we have distinct label sets for each country, while we expect that for the polarity classifier it should be easier to learn useful features across datasets.

In addition to our crosslingual, cross-country MTL models, we evaluate a number of auxiliary tasks that might provide the classifier with relevant information:

**Named Entity Recognition.** Directing the models' attention towards named entities such as persons, organisations etc. might provide useful information for C-type prediction. We use the English Ontonotes dataset (Weischedel et al., 2013) and the German Twitter news headlines data (Ruppenhofer et al., 2020) to train German and English NER models. We use these as auxiliary tasks, in combination with our three subtasks, i.e., C-type prediction for Ireland, Germany and Austria.

| 1 | Fine | Fine | B-PARTY |
|---|---|---|---|
| 2 | Gael | Gael | I-PARTY |
| 3 | health | health | O |
| 4 | spokesman | spokesman | O |
| 5 | Dr | Dr | O |
| 6 | Liam | Liam | B-ACTOR |
| 7 | Twomey | Twomey | I-ACTOR |
| 8 | accused | accuse | O |
| 9 | Fianna | Fianna | B-PARTY |
| 10 | Fáil | Fáil | I-PARTY |

Figure 1: Training data excerpt for auxiliary task (tagging of political actors/coalition terms).

**Political actor tagging.** This auxiliary task aims at injecting some world knowledge about political actors (politicians and parties) and coalitions into the model. As for NER, we use a sequence tagging setup and create our training data as follows. Based on a number of manually created dictionaries, we augment the German input data with BIO labels on the token level (Figure 1). We then use this data in a MTL setting, in combination with C-type prediction for Ireland, Germany and Austria.

**Valence classification.** To improve results for polarity classification, we experiment with a highly related task from the area of sentiment analysis (Zhang et al., 2018). Specifically, we use the multilingual NRC VAD lexicon of Mohammad (2018) which provides valence scores for over 20,000 pivot terms, created via Best-Worst Scaling. We extract the scores for the English words and their German translations and, for each sentence in our German and Irish input data, compute a valence score by simply adding up the scores for each lemma in the sentence and then normalising it by sentence length. We then use the scores in a regression task to make our representations more sensitive towards positive or negative valence in the input signal. The intuition behind this setup is that negative signals should strongly correlate with low valence scores.

| ID | Model | # Epochs | Coalition type | | | C-type + pol | | |
|---|---|---|---|---|---|---|---|---|
| | | | DE | IE | AU | DE | IE | AU |
| | | | **Singletask, monolingual (ST-MO)** | | | | | |
| 1 | MBERT | 20 | 63.8 | 49.0 | 31.6 | 41.2 | 31.2 | 19.7 |
| 2 | MBERT | 25 | 64.0 | 49.9 | 31.6 | 40.8 | 32.9 | 18.3 |
| 3 | MBERT | 30 | 63.0 | 51.4 | 33.1 | 39.6 | 33.9 | 23.0 |
| 4 | R-XLM | 20 | 63.1 | 45.7 | 23.3 | 41.8 | 33.6 | 14.6 |
| 5 | R-XLM | 25 | 64.5 | 47.7 | **36.9** | 42.6 | 34.2 | 24.2 |
| 6 | R-XLM | 30 | 64.1 | 48.0 | 35.9 | 42.0 | 34.0 | 24.0 |
| | | | **Singletask, multilingual (ST-MU)** | | | | | |
| 7 | R-XLM | 20 | 64.5 | 48.5 | 4.8 | 44.5 | 35.8 | 4.7 |
| 8 | R-XLM | 25 | 62.5 | 49.8 | 12.1 | 44.1 | 36.1 | 9.4 |
| 9 | R-XLM | 30 | 65.0 | 49.8 | 12.0 | 44.3 | 36.9 | 6.9 |
| | | | **MTL** | | | | | |
| 10 | MBERT | 20 | 65.0 | 49.4 | 32.0 | 43.1 | 34.2 | 18.1 |
| 11 | MBERT | 25 | 65.5 | 51.7 | 34.7 | 42.4 | 36.1 | **25.8** |
| 12 | MBERT | 30 | **67.3** | **52.3** | 33.3 | 43.3 | 36.4 | 19.5 |
| 13 | R-XLM | 20 | 64.7 | 48.8 | 32.7 | 44.1 | 35.2 | 23.9 |
| 14 | R-XLM | 25 | 63.5 | 50.1 | 34.0 | 43.2 | **38.5** | 25.3 |
| 15 | R-XLM | 30 | 65.5 | 49.7 | 32.3 | 43.9 | 36.9 | 23.7 |
| | | | **MTL (C-type) + ST-MU (pol)** | | | | | |
| 16 | MBERT-30-R-XLM-30 | | **67.3** | **52.3** | 33.3 | **46.7** | **39.4** | 23.1 |

Table 4: Baseline and multitask results for different models (evaluation on signal level)

We compare against the following baselines:

**ST-MO: Singletask, monolingual:** we fine-tune pretrained MBERT (bert-base-multilinual-cased) and XLM-R (xlm-roberta-base) separately for each country and task, resulting in 6 separate classification models.

**ST-MU: Singletask, multilingual:** we merge all training data into one set and use the multilingual XLM-R to train two unified models, one for polarity and one for predicting C-types (this approach crucially increases the number of country-specific C-types in the data while the number of polarity labels remains the same).

### 5.2 Multitask learning results

Table 4 reports F1 for C-type prediction and for the combined final labels, C-type+polarity for our different experimental setups. While our datasets are not small (>27,000 sentences), due to the sparsity of the phenomenon and fineness-of-grain of our C-type schema we are still facing a sparse data problem. We therefore opted for reporting results over different epochs, to show the variation in results which seemed more informative than a single score. When applying the models to new data, how-ever, we recommend the use of an ensemble system (e.g., a simple majority vote over the predictions of different models) to avoid selecting an inferior model based on a non-representative dev set.

**Singletask, monolingual (ST-MO, Exp. 1-6).** Looking at experiments 1-6, we notice that for German, MBERT results are slightly lower than the ones reported in § 4.4 (see Table 3), especially for the final, atomic labels (C-type+pol). This is to be expected, given that our previous experiments were using the BERT model specifically trained for German. As expected, results for Ireland and Austria are substantially lower than the ones for Germany, reflecting the smaller sizes of the respective training sets. This is a problem especially for the Austrian data with less than 3,000 sentences.

**Singletask, multilingual (ST-MU, Exp.7-9).** Experiments 7-9 show that for C-type prediction, jointly training the model on the union of all three datasets yields a slight improvement for the largest dataset (DE). For Ireland, results are in the same range as before while for the smallest dataset (AU), we observe a drastic decrease in F1. German and Irish F1 scores for C-type+polarity, however, are around 2% higher than for ST-MO, thus showing

| Model | DE | IE | AU |
|---|---|---|---|
| **Best C-type (Exp.12)** | **67.3** | **52.3** | **33.3** |
| MTL + NER | | | |
| R-XLM-30 | 64.5 | 48.5 | 4.8 |
| MBERT-30 | 66.2 | 52.0 | 30.1 |
| MTL + political actors | | | |
| MBERT-30 | 64.3 | 50.0 | 31.8 |

Table 5: MTL results for coalition type prediction for auxiliary tasks (evaluation on signal level)

| Model | DE | IE | AU |
|---|---|---|---|
| **Best C-type+pol (Exp.16)** | 46.7 | 39.4 | 23.1 |
| MTL + valence | | | |
| R-XLM-25 | 45.8 | 39.9 | 23.2 |
| R-XLM-30 | **47.6** | **41.2** | **24.5** |

Table 6: MTL results for polarity prediction for auxiliary tasks (evaluation on signal level)

the improvement of the ST-MU polarity classifier over the monolingual ones. Again, this makes sense, considering that – in contrast to the C-types – the polarity labels are shared across countries.

**MTL results (Exp.10-15).** We now turn to experiments 10-15 where we jointly train a shared representation for each subtask by considering each country as a separate task in our MTL setup. For both subtasks, C-type prediction and polarity classification, we observe some improvements over the ST baselines. Results for German and Irish C-types increase by more than 2% F1 (Exp.12) over the best ST results, and for polarity we see improvements for Ireland and Austria (but not for Germany).[10]

Our modular setup allows us to combine the best classifiers from each setting. As we noticed that for polarity prediction, simply merging all three datasets into one training set seemed to work well, due to the unified labels across countries and the larger training size, we combine the best C-type MTL setup (Exp.12: MBERT-30) with the best polarity model (Exp.9: R-XLM-30). Results for Ireland and Germany (Exp.16) now outperform the ST baselines (ST-MO, Exp.1-6) by 4-5% F1.

### 5.3 Results for auxiliary tasks

Results for the auxiliary tasks created to **improve C-type prediction** (NER and political actor tagging, Table 5) fail to surpass our best MTL results. A possible explanation is 'catastrophic interference' (Hashimoto et al., 2017) where learned information is overwritten by the new tasks.

The auxiliary task designed to **improve polarity prediction** (Table 6) via learning to predict valence scores, on the other hand, achieves further improvements in the range of 1-2% over previous scores and results in best scores for the combined C-type+polarity labels for all three countries.[11]

---

## 6 Conclusions and Future Work

We presented experiments on coalition signal detection, a new, challenging NLU task for the political text analysis domain, and showed that a simple transfer learning approach outperforms previous work based on dictionary-based methods.

We further showed that coalition signal detection in a multilingual setting is feasible and can substantially outperform results for the monolingual setup. We also approached the problem in a multitask learning setup, modelling the prediction of inter-party communication for different countries as separate tasks. This approach gave us improvements over the monolingual singletask setting. We could further increase results by adding another, highly related auxiliary task (valence prediction), obtaining final improvements of 5-7 percentage points (DE, IE) over the monolingual baselines.

For future work, we would like to test adapters (Pfeiffer et al., 2020) as another way to inject relevant information into the model and mitigate the issue of 'catastrophic forgetting' that might be responsible for the negative results we obtained for two auxiliary tasks (NER, political actor tagging).

## 7 Ethics statement

The intended use of this work is to enable researchers to conduct large-scale longitudinal studies that might increase our understanding of the role of coalition signals in electoral campaigns, while avoiding the high time requirements and cost for manual coding. While our results show some progress over previous work, there are a number of pitfalls that we would like to point out. First, our experiments showed that identifying coalition signals in newswire is a very challenging task and that results are not yet very accurate. We would thus like to stress the need for additional external validation of the results. Another pitfall of applying machine learning to coalition signal detection for

---

new elections comes from the fact that we might observe shifts in the distribution of signals in new, unseen data. New parties and coalition types might emerge or previously frequent signals might become less frequent. This is a problem for supervised machine learning as classifiers tend to learn the distribution in the training data and thus might introduce an unwanted bias that does not fit the distribution of new elections. We address this issue in subsequent work by creating synthetic training instances for sparse coalition types in the training data and show that this approach can ameliorate the problem (Adendorf et al., 2021).

## Acknowledgements

## References

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

Gavin Abercrombie, 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.

Anna Adendorf, Ines Rehbein, Oke Bahnsen, Thomas Gschwend, Simone Paolo Ponzetto, and Lukas F. Stoetzer. 2021. Who wants to go with whom? Identifying coalition signals in newspaper articles using transfer learning. *Working Paper*.

Zachary J Auter and Jeffrey A Fine. 2016. Negative campaigning in the social media age: Attack advertising on facebook. *Political Behavior*, 38(4):999–1020.

Oke Bahnsen, Thomas Gschwend, and Lukas F. Stoetzer. 2020. How do coalition signals shape voting behavior? Revealing the mediating role of coalition expectations. *Electoral Studies*, 66(May):102166.

Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774.

Volker Best. 2015. *Koalitionssignale bei Landtagswahlen : Eine empirische Analyse von 1990 bis 2012*, 1. auflage edition. Parteien und Wahlen; Band 9. Nomos, Baden-Baden.

Nico Blokker, Erenay Dayanik, Gabriella Lapesa, and Sebastian Padó. 2020. Swimming with the tide? positional claim detection across political text types. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 24–34, Online. Association for Computational Linguistics.

Shaun Bowler, Gail McElroy, and Stefan Müller. 2021. Voter expectations of government formation in coalition systems: The importance of the information context. *European Journal of Political Research*.

Richard Caruana. 1993. Multitask learning: A knowledge-based source of inductive bias. In *The Tenth International Conference on Machine Learning*, pages 41–48. Morgan Kaufmann.

Luigi Curini. 2011. Negative campaigning in no-cabinet alternation systems: Ideological closeness and blames of corruption in italy and japan using party manifesto data. *Japanese Journal of Political Science*, 12(3):399–420.

Luigi Curini and Paolo Martelli. 2010. Ideological proximity and valence competition. negative campaigning through allegation of corruption in the italian legislative arena from 1946 to 1994. *Electoral Studies*, 29(4):636–647.

David F Damore. 2002. Candidate strategy and the decision to go negative. *Political Research Quarterly*, 55(3):669–685.

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.

Wouter De Nooy and Jan Kleinnijenhuis. 2013. Polarization in the media during an election campaign: A dynamic network nodel predicting support and attack among political actors. *Political Communication*, 30(1):117–138.

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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Martin Dolezal, Laurenz Ennser-Jedenastik, and Wolfgang C Müller. 2016. Negative campaigning and the logic of retaliation in multiparty competition. *The International Journal of Press/Politics*, 21(2):253–272.

Martin Dolezal, Laurenz Ennser-Jedenastik, and Wolfgang C Müller. 2017. Who will attack the competitors? how political parties resolve strategic and collective action dilemmas in negative campaigning. *Party Politics*, 23(6):666–679.

Martin Dolezal, Laurenz Ennser-Jedenastik, Wolfgang C Müller, Katrin Praprotnik, and Anna Katharina Winkler. 2018. Beyond salience and position taking: How political parties communicate through their manifestos. *Party Politics*, 24(3):240–252.

James N Druckman, Martin J Kifer, and Michael Parkin. 2009. Campaign communications in u.s. congressional elections. *American Political Science Review*, 103(3):343–366.

Christian Elmelund-Præstekær. 2008. Negative campaigning in a multiparty system. *Representation*, 44(1):27–39.

Christian Elmelund-Præstekær. 2010. Beyond american negativity: Toward a general understanding of the determinants of negative campaigning. *European Political Science Review*, 2(1):137–156.

Christian Elmelund-Præstekær. 2011. Issue ownership as a determinant of negative campaigning. *International Political Science Review*, 32(2):209–221.

Heather K Evans, Victoria Cordova, and Savannah Sipole. 2014. Twitter style: An analysis of how house candidates used twitter in their 2012 campaigns. *PS Political Science & Politics*, 47(2):454–462.

Bryan T Gervais, Heather K Evans, and Annelise Russell. 2020. Fear and loathing on twitter: Exploring negative rhetoric in tweets during the 2018 midterm election. In Sean D Foreman, Marcia L Godwin, and Walter C Wilson, editors, *The Roads to Congress 2018*, pages 31–52. Springer.

Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. 2019. Computational analysis of political texts: Bridging research efforts across communities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 18–23, Florence, Italy. Association for Computational Linguistics.

Justin Grimmer. 2010. A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis*, 18(1):1–35.

Justin Grimmer and Brandon M. Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297.

Justin H Gross and Kaylee T Johnson. 2016. Twitter taunts and tirades: Negative campaigning in the age of trump. *PS Political Science & Politics*, 49(4):748–754.

Thomas Gschwend, Michael F. Meffert, and Lukas F. Stoetzer. 2017. Weighting Parties and Coalitions: How Coalition Signals Influence Voting Behavior. *Journal of Politics*, 79(2):642–655.

Martin Haselmayer and Marcelo Jenny. 2018. Friendly fire? negative campaigning among coalition partners. *Research & Politics*, 5(3):1–9.

Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher. 2017. A joint many-task model: Growing a neural network for multiple NLP tasks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1923–1933, Copenhagen, Denmark. Association for Computational Linguistics.

Richard R Lau and Gerald M Pomper. 2002. Effectiveness of negative campaigning in u.s. senate elections. *American Journal of Political Science*, 46(1):47–66.

Will Lowe, Kenneth Benoit, Slava Mikhaylov, and Michael Laver. 2011. Scaling Policy Preferences from Coded Political Texts. *Legislative Studies Quarterly*, 36(1):123–155.

Stefano Menini, Federico Nanni, Simone Paolo Ponzetto, and Sara Tonelli. 2017. Topic-based agreement and disagreement in US electoral manifestos. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2938–2944, Copenhagen, Denmark. Association for Computational Linguistics.

Saif M. Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *NIPS Autodiff Workshop*.

Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2020. AdapterHub: A framework for adapting transformers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 46–54, Online. Association for Computational Linguistics.

Sebastian Ruder. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Josef Ruppenhofer, Ines Rehbein, and Carolina Flinz. 2020. Fine-grained named entity annotations for German biographic interviews. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4605–4614, Marseille, France. European Language Resources Association.

Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. 2020. GPolS: A contextual graph-based language model for analyzing parliamentary debates and political cohesion. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4847–4859, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Lukas F. Stoetzer and Matthias Orlowski. 2020. Estimating coalition majorities during political campaigns based on pre-election polls. *Journal of Elections, Public Opinion and Parties*, 30(1):126–137.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Annemarie S Walter and Wouter Van der Brug. 2013. When the gloves come off: Inter-party variation in negative campaigning in dutch elections, 1981–2010. *Acta Politica*, 48(4):367–388.

Annemarie S Walter and Cees Van der Eijk. 2019. Measures of campaign negativity: Comparing approaches and eliminating partisan bias. *The International Journal of Press/Politics*, 24(3):363–382.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, and Ann Houston. 2013. OntoNotes Release 5.0.

Siegfried Weischenberg, Maja Malik, and Armin Scholl. 2006. Journalismus in Deutschland 2005. *Media Perspektiven*, 7:346–361.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020*, pages 38–45. Association for Computational Linguistics.

Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis : A survey.

## A   Appendix

| ID | Germany | # | Ireland | # | Austria | # |
|---|---|---|---|---|---|---|
| 1 | CDU + FDP | 345 | Fianna Fáil + Fine Gael | 160 | SPÖ + ÖVP | 38 |
| 2 | CDU + SPD | 338 | Fine Gael + Labour | 124 | FPÖ + SPÖ | 17 |
| 3 | Greens + SPD | 172 | Fianna Fáil + Sinn Féin | 74 | BZÖ + SPÖ | 11 |
| 4 | FDP + Greens + SPD | 172 | Fianna Fáil + Labour | 64 | FPÖ + ÖVP | 10 |
| 5 | CDU + FDP + Greens | 83 | Fianna Fáil + PD | 45 | Greens + ÖVP | 7 |
| 6 | Left + SPD | 73 | Fianna Fáil + Greens | 29 | BZÖ + FPÖ + ÖVP | 5 |
| 7 | Greens + Left + SPD | 66 | Fine Gael + Greens | 21 | FPÖ + Team Stronach + ÖVP | 5 |
| 8 | CDU + Greens | 64 | Fine Gael + Sinn Féin | 11 | NEOS + SPÖ + ÖVP | 4 |
| 9 | CDU + other | 23 | Independent + unspecified | 8 | | |
| 10 | FDP + SPD | 22 | Fine Gael + Greens + Labour | 8 | | |
| 11 | SPD + unspecified | 18 | Labour + Sinn Féin | 7 | | |
| 12 | FDP + unspecified | 13 | Greens + Labour | 5 | | |
| 13 | CDU + CDU (CDU/CSU) | 10 | | | | |
| 14 | Greens + Left | 10 | | | | |

Table 7: List of coalition types (C-types) used for training and their distribution (#) in the data (CDU: Christian German Union, FDP: Free Democratic Party, SPD: Social Democratic Party, PD: Progressive Democrats, BZÖ: Alliance for the Future of Austria, FPÖ: Freedom Party of Austria, NEOS: The New Austria and Liberal Forum, ÖVP: Austrian People's Party, SPÖ: Social Democratic Party of Austria).

| ID | Germany | F1 | Ireland | F1 | Austria | F1 |
|---|---|---|---|---|---|---|
| 1 | CDU + FDP | 68.4 | Fianna Fáil + Fine Gael | 65.9 | SPÖ + ÖVP | 63.8 |
| 2 | CDU + SPD | 78.5 | Fine Gael + Labour | 72.6 | FPÖ + SPÖ | 36.8 |
| 3 | Greens + SPD | 67.4 | Fianna Fáil + Sinn Féin | 71.6 | BZÖ + SPÖ | 16.7 |
| 4 | FDP + Greens + SPD | 86.5 | Fianna Fáil + Labour | 51.8 | FPÖ + ÖVP | 0.0 |
| 5 | CDU + FDP + Greens | 70.4 | Fianna Fáil + PD | 65.6 | Greens + ÖVP | 28.6 |
| 6 | Left + SPD | 60.3 | Fianna Fáil + Greens | 51.3 | BZÖ + FPÖ + ÖVP | 0.0 |
| 7 | Greens + Left + SPD | 54.0 | Fine Gael + Greens | 45.4 | FPÖ + Team Stronach + ÖVP | 40.0 |
| 8 | CDU + Greens | 50.9 | Fine Gael + Sinn Féin | 20.0 | NEOS + SPÖ + ÖVP | 0.0 |
| 9 | CDU + other | 84.6 | Independent + unspecified | 30.8 | | |
| 10 | FDP + SPD | 10.0 | Fine Gael + Greens + Labour | 36.4 | | |
| 11 | SPD + unspecified | 0.0 | Labour + Sinn Féin | 18.2 | | |
| 12 | FDP + unspecified | 0.0 | Greens + Labour | 0.0 | | |
| 13 | CDU + CDU (CDU/CSU) | 0.0 | | | | |
| 14 | Greens + Left | 0.0 | | | | |

Table 8: F1 for model 12 from Table 4 for each C-type in the data (evaluation on the signal level), excluding polarity (CDU: Christian German Union, FDP: Free Democratic Party, SPD: Social Democratic Party, PD: Progressive Democrats, BZÖ: Alliance for the Future of Austria, FPÖ: Freedom Party of Austria, NEOS: The New Austria and Liberal Forum, ÖVP: Austrian People's Party, SPÖ: Social Democratic Party of Austria).

| ID | Germany | pos | neut | neg | Ireland | pos | neut | neg |
|---|---|---|---|---|---|---|---|---|
| 1 | CDU + FDP | 65.1 | 11.8 | 11.8 | Fianna Fáil + Fine Gael | 43.2 | 53.7 | 53.3 |
| 2 | CDU + SPD | 52.2 | 41.9 | 37.0 | Fine Gael + Labour | 69.2 | 22.2 | 31.6 |
| 3 | Greens + SPD | 57.1 | 20.0 | 0.0 | Fianna Fáil + Sinn Féin | 26.1 | 18.2 | 61.1 |
| 4 | FDP + Greens + SPD | 33.3 | 42.1 | 71.1 | Fianna Fáil + Labour | 40.0 | 33.3 | 37.0 |
| 5 | CDU + FDP + Greens | 26.7 | 23.5 | 58.5 | Fianna Fáil + PD | 63.4 | 33.3 | 16.7 |
| 6 | Left + SPD | 33.3 | 66.7 | 66.7 | Fianna Fáil + Greens | 40.0 | 0.0 | 66.7 |
| 7 | Greens + Left + SPD | 42.1 | 40.0 | 28.6 | Fine Gael + Greens | 52.6 | 0.0 | 0.0 |
| 8 | CDU + Greens | 40.0 | 35.3 | 16.7 | Fine Gael + Sinn Féin | 0.0 | n.a. | 23.1 |
| 9 | CDU + other | 0.0 | 0.0 | 90.9 | Independent + unspecified | 0.0 | 0.0 | 66.7 |
| 10 | FDP + SPD | 0.0 | 0.0 | 25.0 | Fine Gael + Greens + Labour | 22.2 | n.a. | 0.0 |
| 11 | SPD + unspecified | 0.0 | 0.0 | 0.0 | Labour + Sinn Féin | n.a. | n.a. | 18.2 |
| 12 | FDP + unspecified | 0.0 | 0.0 | n.a | Greens + Labour | 0.0 | n.a. | n.a. |
| 13 | CDU + CDU (CDU/CSU) | 0.0 | 0.0 | 0.0 | | | | |
| 14 | Greens + Left | 0.0 | n.a. | 0.0 | | | | |

Table 9: F1 for the best performing model from Table 6 (R-XLM-30) for each C-type + polarity in the German and Irish data (evaluation on the signal level) (CDU: Christian German Union, FDP: Free Democratic Party, SPD: Social Democratic Party, PD: Progressive Democrats).