Partisan Semantic Overlaps: Floor-speeches and Ideological Position

Benjamin Guinaudeau[1]
Simon Roth
University of Konstanz

October 25, 2020[2]

Abstract

Estimating the ideological position of Members of Parliaments (MPs) remains a challenge for political scientists. Different approaches have been developed including surveys, roll-call votes and floor speeches. Inspired by the measure of polarization proposed in Peterson and Spirling (2018), we present a new unsupervised strategy to extract ideological positions from speeches. We rely on partisan semantic overlaps (PSO), defined as language patterns indistinguishably used across parties. We train artificial neural networks to predict party labels given text and expect these semantic overlaps to be mapped by the partisan probabilities. The higher the overlap between two MPs, the smaller is their ideological distance.

We use three decades of parliamentary speeches in six countries (Canada, France, Germany, New Zealand, Spain, United Kingdom) and estimate, in each of these countries, partisan probabilities with a convolutional network. We show party-level positions are accurately captured by the measure (high correlation with CMP). In the absence of any broadly accepted individual ideological measure, we use a new expert survey designed to capture MPs' position to validate our ideological scores at the individual level. For now, intra-partisan heterogeneity is not accurately captured. We discuss the potential origins of these results and propose possible ways to address these in the future.

[1] *benjamin.guinaudeau@uni-konstanz.de*
[2] This draft is work in progress. Please do not circulate it without the permission of the authors.

1 Introduction

Measuring ideology has always been a big challenge for political scientists. Many efforts were invested in deploying a whole array of strategies aiming at the same target: ideology and its many forms. The position of both voters and collective groups were measured using surveys. Models analyzing these surveys became increasingly complex and, most importantly, more accurate. Political actors were not spared and were also subjected to ideological measures. Facing the need of overcoming survey approaches, contributions became more creative and/or more technical. They relied alternatively on roll-call votes, speeches, manifestos, social-media posts and networks, lobby ratings, etc. Thanks to these efforts, we certainly understand better ideology and its role in politics. Nevertheless, no measurement strategy was crowned as the valid way to measure ideology across context. But, if all these efforts failed to deliver a golden standard strategy, able to accurately and automatically measure ideology in any setting, this is probably because such measure does not exist.

The concept of ideology is ubiquitous and has broadly been broadly investigated. However, its empirical operationalisation remains a diffuse and moving target. Ideologies are continuously evolving. It is hence crucial for political scientists to keep undertaking two different tasks. First, we must review existing strategies to ensure they still deliver accurate estimates. This review process is especially important for inductive measures, whose validity relies on a strict definition of the ideological dimension. Second, we need to keep innovating by developing new strategies, which incorporate progress made in other fields. Combining the review of previous strategies with the discovery of new ones will help us, even if none is perfect, to approximate the best possible measure. This paper aims at proposing a new deductive - no meaning of the ideological dimension is set ex-ante - strategy to measure ideology based on texts. While previous text-based measures modelled language and approach ideology as a latent factor influencing the relationship between an actor and its word choice, the measure presented in this paper relies on, what we call, partisan semantic overlaps. Put briefly, we use convolutional neural networks trained to identify the party label of a speaker given one her speech. The model is overfitted, so that the remaining uncertainty is caused by wording or phrasing choices common to several parties. The uncertainty regarding the belonging a text to parties defines partisan semantic overlaps. We expect these overlaps to be structured along the most salient ideological dimension, so that close ideological position exhibit similar partisan probability structures.

The rest of the paper proceeds as follows. After a brief discussion of the concept of ideology, we present a typology of the existing measures, organized along the criteria used by measurement strategies to identify the left-right dimension. We distinguish between survey, deductive behavioural and inductive behavioural measures. Because previous attempts to measure ideology based on speech are very close to the present paper, we then discuss this specific approach more deeply. Building on the experience of these speech-based measures, we finally partisan semantic overlaps. To illustrate the approach, the measure is implemented on the speeches from six different parliaments. With the results, we present evidence of face-validity and convergence validity (but only at the partisan level). To conclude, we discuss potential issues, which might drive the current unsatisfying results.

2 Measuring Ideology

2.1 Summarizing Complex Policy Preferences

We understand political ideology as a coherent system linking together political opinions (Carmines and D'Amico 2015). As policy preferences are high-dimensional and complex, ideology proves to be a necessary tool for any political system to work, providing important heuristics. Crucial political phenomena such as elections, representation, political parties and even the simplest political discussion needs quick and stable way to summarize complex political opinions. Without ideology, difficult-enough task such as voting or building a policy coalition would be impossible. Political actors have, thus, naturally developed ideological cues, which are nothing else but a process of reducing the dimensionality of policy preferences to one or two dimensions. Measuring ideology of political actors amounts to finding the recipe of this dimensionality reduction.

Unfortunately, this system of rules linking individual preferences to ideological position is not exogenously set. As they observe and take part to the political routine, actors discover and collectively shape this recipe. Luckily, ideology needs to be stable to fulfill its purpose, so the link between political preferences and ideological position is not completely reinvented every day. It does, however, evolve slowly and, sometimes, it even undergoes some deep and sudden realignment. For instance, as the salience of political issues goes up and down, their link with the main ideological dimension varies as well. This evolution makes the measure harder and continuously renew the challenge of measuring and interpreting ideology.

2.2 Ideology is a Moving Target

Ideology plays a constitutive role in most political processes and measuring it accurately benefits any subfields of political science. The benefits even spillover to other disciplines, such as economics, sociology or history. This explains the large number contributions dedicated to this task. The trade-off between cross-context comparability and context adaptability is one example of the challenges faced by ideological measures. Because ideology is continuously and collectively redefined, it does not hold the same meaning in different contexts (Benoit and Laver 2006). A left-leaning position in the USA does not amount to the same policy preferences as a left-leaning position in Germany. This also works over time. The issues characterizing right-leaning position in any country certainly changed over the past fifty years. When measuring ideology, we attempt to systematically associate a numeric value with a position. But it is impossible to represent the variety of ideological recipes, without losing precision or accuracy. If we were interested in comparing German and American ideological positions, we could identify an ideological scale that is approximately accurate in both contexts. With this scale, a German ideological score of 3 could be compared with an American score of 3. But this would sacrifice precision, considering that a score of 3 on the overall scale might in fact be a 2 in a purely German scale. The alternative strategy would be to enhance contextual precision at the cost of comparability. We could, as in the above-mentioned example, use the German scale instead of the overall scale and accurately locate German actors. But, then, a German 2 would not be comparable anymore with an American 2. Examples for both types of measure exist. The Comparative Manifesto Project (CMP) (Volkens et al. 2017) chooses to use the same definition of left and right in all countries. In contrast, the Chapel-Hill Expert Survey (CHES) (Hooghe et al. 2010) asks national experts to estimate the ideological position of parties given the national context. Both measures are heavily used in the comparative literature.

The trade-off between adaptability and comparability - also known as concept stretching - is well identified in political science. Since the seminal work of Sartori (1970), it has been extensively discussed in many different contexts. It offers one example of the reasons making the puzzle of measuring ideology homogeneously unbreakable. If no strategy can measure ideology in any circumstances, this measurement issue can still be tackled by a diversity of approaches, whose individual superiority depends on the research question and on the empirical context.

2.3 Existing strategies

Dozens of contributions have claimed to measure ideology. We do not attempt to review each of them and inspired by Laver (2014), we summarize them, instead, through a typology, based on the criteria identifying the substantial meaning of the main ideological dimension. Three types of measure are distinguished survey, deductive behavioural and inductive behavioural.

2.3.1 Survey Measures

Survey measures do not require any definition of left and right. They assume these concepts to be known to the respondent and rely on their interpretation. Usually, they ask participants to estimate the position of known actors - usually themselves, parties or political leaders- on an eleven-scale going from 0 to 10. This has been mostly used in the context of international large-n individual surveys - Comparative Survey of Electoral Systems, Eurobarometer, etc.- and expert surveys (CHES). Survey measures rely on the respondents' subjective understanding of what is left and right. In the aggregate, this understanding should be accurate, but adapted to each context. The assumption that an ideological score of 3 amounts to the same position in France and Germany can hardly be defended (Kim and Fording 1998). This heterogeneity can, however, be statistically modelled (Caughey, O'Grady, and Warshaw 2019).

Survey measures come with three shortcomings: they are expensive, have a limited precision and they are limited to a certain type of actors. Pricing issue is self-explaining: running a survey costs money. The precision of the resulting measure is constrained by the scale used in the question. It seems common to use an eleven-point scale, so that they *only* eleven different ideological position can in the end be distinguished. But the ideological reality is much more complex and many research questions require to represent ideology as a continuum. The simplification of the scale is necessary to reduce the risk of differing scale interpretations across the subjects. If respondents certainly agree on what is left and right, they are less likely to agree on what represents a score of 3 on an eleven-point scale and will almost certainly disagree on what distinguishes a score 31 from a 32 on a 100-point scale. Reducing the risk of diverging interpretations also reduces the precision of the measure. In addition, survey measures can only be deployed for self-estimation or the estimation of a limited number of political actors and organizations. It is hardly conceivable to think of any expert with enough knowledge to accurately estimate the ideological position of all individual members of parliament (MPs). Consequently, survey measures constitute a good alternative if money is available, if the required estimate does not need to be fine-grained and if the targeted actors are either the respondents or a small number of prominent political actors.

Behavioural measures constitute the main alternative to survey. They rely on the assumption that ideological positions translate systematically into observable behavioural patterns. The idea is hence to deduce from an observed behaviour, such as delivered speeches or votes, the underlying ideological position. Behavioural measures mobilize two different strategies to identify the left-right dimension. They proceed either deductively or inductively (Laver 2014).

2.3.2 Inductive Behavioural Measures

Inductive measures specify ex-ante the meaning of left and right. Some of them hard-code the behaviour, others select behaviours that can be used as anchors to scale out-of-sample behaviours.

The well-established Comparative Manifesto Project (Volkens et al. 2017) is an example of the former approach. They intend to measure the ideological position of parties based on their electoral manifestos. To do so, the team of the CMP attributes to each quasi-sentence a label that encodes both the topic and the defended position. Some of these labels are predefined as left-leaning - e.g. less military ; market regulation or welfare expansion - others as right-leaning - e.g. more military ; free market or welfare limitation-. Their coding scheme makes it possible to compute the left-leaning and right-leaning proportions of manifestos, whose difference is assimilated as the position of the party on the left-right dimension.

Wordscores (Laver, Benoit, and Garry 2003 ; Lowe 2008) is another behavioural strategy that inductively identify the left-right dimension. It is a text-based method, whose objective is to measure the ideological positions of a corpus of virgin documents. It starts with the identification of anchor documents. These anchors need to represent typical far-left and far-right speeches. In a second step, using words distributions, Wordscores computes the ideological loading of each word and estimates the ideological distance between the virgin and the anchor documents.

Inductive behavioural measures raise two types of interrogations. First, the validity of predefined left and right can be questioned. This relates, among others, to the comparability-adaptability trade-off mentioned earlier. The resulting measure can only be as valid as the anchors. The measure proposed by the CMP uses an international and overall left-right definition. If the resulting estimates are comparable across countries, they might suffer from contextual measurement errors. The validity of Wordscores estimate is very sensitive to the chosen anchors. For the measure to be valid, the anchors must match the measurement context. In fact, choosing valid and informative anchor texts proved to be a difficult task (Proksch and Slapin 2009). The second issue regards the estimation of the numeric values. For CMP, a larger proportion of right-wing topics amounts to a more right-leaning position. But this is not necessarily the case. Centre-right parties might spend a large proportion of their manifestos talking about right-wing topics, without being far right. This issue is even more salient considering the heterogeneity of manifestos across parties. This second group of measures is accordingly an interesting approach, which mixes human judgment with automatic scaling methods. When deployed, two points should be carefully scrutinized: the adequacy between the ex-ante definition of left and right and the validity of the computation translating the behavioural distances into numeric values.

2.3.3 Deductive Behavioural Measures

Deductive measures discover automatically the ideological dimension in the data. These data-driven approaches rely consequently on very strong theoretical assumptions regarding the data generating process. While inductive measures capture the distance between typical and targeted behaviour, deductive one estimate the dominant dimension structuring behaviours and assimilate it with the left-right dimension.

For example, roll-call analyses (poole2000congress ; Poole 2005) assume the individual voting behaviour of MPs to map the dominant ideological dimension. Right-leaning legislators are expected to vote together and differently from their left-leaning colleagues. Deductive measures become, however, problematic, when the observed behaviour is affected by other factors. Members of government parties usually tend to vote together notwithstanding their ideological position. This produced estimates mapping the government-opposition divide instead of the left-right divide. Additionally, the observed behaviours need to be representative of the overall ideological position. Again, in the case of roll-call votes, it has been shown that this voting procedure is not genuinely triggered and responds to important selection mechanisms. This systematic bias prevents from extrapolating the general position of an MP based on her observed roll-call behaviour. Finally, the observed behaviour needs to be informative enough to distinguish between the different ideological positions motivating it. Both far-right and far-left legislators often oppose legislation and vote together, even if their opposition is motivated by very different reasons. If this happens, roll-call estimates would cluster these two parties together, even though they hold very different positions. In addition, partisan discipline drastically limit the observed variance. If legislators do not defect from their party line, it is impossible to identify members voting with more left-leaning parties from members voting with right-leaning parties.

So far, we focused only on roll-call analyses, as an example of inductive method. Other inductive measures have exploited parliamentary speeches. This group of strategies - including partisan semantic overlap - rely on the assumption that the wording and the phrasing of floor speeches are primarily driven by the ideological position of an MP. We discuss in more details these approaches later in this paper. To conclude on deductive approaches, their validity lie in the strong assumption, that the dominant dimension structuring behaviour amounts to the left-right dimension. But many things can go wrong and this assumption has often good reasons to be violated.

It is therefore even more important to properly validate ex-post the estimates.

Overall, behavioural measures constitute important attempts to measure ideology with data, that is often publicly available. In that sense, they are less expensive than surveys and their scope is only limited by the data availability. Since they produce continuously scaled estimates, the do not suffer from limited precision. This being said, the centrality of their theoretical assumptions, combined with the usual complexity of the involved computation, calls for a careful and proper validation. The absence of golden standard makes the validation more difficult. In the past authors tested for face and convergence validity. In the absence of previous estimate - as it is the case with MPs -, it is, however, not enough to validate the measure at the partisan level. Because invalid individual estimates might produce valid aggregated ones, individual measures need to be validated at the individual level.

This section distinguished between three different groups of ideological measures depending on how they identify the left-right dimension. The strategy presented in this paper is a deductive behavioural measure and takes advantages of floor speeches. In contrast to existing text-scaling methods, which relies on language modelling assumptions, we propose here to identify the left-right dimension using partisan semantic overlaps.

2.4 Speech-Based Measures of Ideology

2.4.1 Language Models

Three different deductive ideological measures were already proposed. They all use a similar approach and built on each other, so that each new measure addressed limitations of the previous one. In 2008, Slapin and Proksch (2008) developed Wordfish, assuming that the word count distributions of political documents - they initially used party manifestos, but applied later the method to parliamentary speeches - follow a Poisson distribution, whose mean depends on the ideological position of the author. More precisely, to scale the position of parties, they include party fixed effect in a Poisson regression and assume the corresponding fixed effect to be structured along the left-right dimension. Eight years later, Lauderdale and Herzog (2016) proposed Wordshoal. They criticize Wordfish for its inability to accurately capture the left-right dimension if the documents are not about topics, structuring the main ideological dimension. To solve this issue, they split the corpus into topical debates, estimate debate-specific parameters and map these parameters onto one single latent dimension using a Bayesian linear factor mode. They claim that using Wordfish on documents dealing with the same topic allows capturing preferences instead of a mixture of topics. Finally, Rheault and Cochrane (2020) takes advantage of recent natural-language processing techniques to overcome the bag-of-word representation as used in Wordfish and Wordshoal. When representing text as a document-term-matrix (word counts), the word is not interpreted in its context. Now, certain words hold different ideological loadings depending on the context. Consider the two following sentences. “The security of our border should be a priority” and “Social security should be a priority”. The word *security* is used in each of these sentences, but because of the context informs about different ideologies in the two cases. While n-gram can help to take context into account, this quickly increases both the dimensionality of the data and its sparsity, which makes modelling even more complicated. Instead, they use Word-Embedding, which is a common strategy to represent text in a way that takes context into account (Mikolov et al. 2013). They use a neural network with an embedding layer to predict a word given its context and add partisan or legislator fixed effect depending on their target. In addition, they control for confounding factors, such as government, by including a dummy for government parties. They extract the principal component of these fixed effects and use them as ideological estimates.

Put briefly, each of these strategies model language (word probability or word count) as a function of the ideological position of the author and expect the relationship between targets - party or MP - and word choice to be structured along the main ideological dimension. Wordfish estimates the count of a given word given the party as expressed in the following equation. In doing so, they obtain for each party, a vector β, whose principal component is extracted using PCA.

$$Count(word_{wi}) = Poisson(\alpha + \beta_k * party_{ki} + \beta * X_i)$$

Rheault and Cochrane (2020) estimates a similar model, which as presented in the following equation, where h is the neural net, $context_i$ the words surrounding the targeted word and gov_j a dummy measuring whether the speaker belongs to a governing party. As for Wordfish, they obtain a vector of coefficient, which is mapped onto one dimension using PCA.

$$Word_{ji} = h(\alpha + \beta_j * speaker_{ji} + \beta_1 * context_i + \beta_2 * gov_j)$$

Because it also uses neural networks, our strategy is in many ways very similar to this last contribution. The left-right identification is, however, completely different and relies on partisan semantic overlaps. Instead of modelling language, we use language to identify similarities between parties and expect the dominant dimension of these similarities to map the left-right dimension.

2.4.2 Partisan Semantic Overlaps

The present measure relies on the central assumption that the language patterns shared between parties are structured along the left-right dimension. If a specific speech can be equally attributed to two parties, we assume it informs about an ideological position, which is common to these two parties. We call those shared language patterns *partisan semantic overlaps*. Semantic overlapping between parties is not necessarily caused by ideological proximity. Sentences criticizing the government might, for instance, be used by any of the opposition parties, regardless of their ideology. If not fitted correctly, semantic similarities can capture mixtures of topics, instead of ideological positions. We try to address both of these issues when modelling the overlaps.

We propose to use floor speeches to estimate the partisan semantic overlaps. Using a neural network, we estimate for each text its probability to belong to any of the parties. The neural network is composed out of embedding, convolutional and LSTM - Long Short-Term Memory- layers. The embedding layer transforms text from a sequence of discrete tokens to a dense vector representation. Subsequently, a 1D convolutional layer extracts n-gram features that are fed into an LSTM layer, which connects those low-level features over time. All layers are optimized according to our global loss function minimizing the error of the predicted party labels. To distinguish between partisan overlaps and institutional overlaps caused by the government-opposition divide, we also predict whether the author belongs to a government party. Using a softmax activation function, we ensure that the model identifies a speech as either typical of a party or typical of the government. The combination of neural networks with sequential text representation allows removing any preprocessing step. The estimation efficiency of neural networks allows using a huge vocabulary (15 000 unique tokens), without extending indefinitely the duration of the estimation. For this exact reason, stemming or lemmatising are not necessary to increase the amount of information provided to the model. For any $text_i$, we estimate the probability of this text to belong to each of the k parties and its probability to belong to the government. The output layer has accordingly $k + 1$ dimensions.

$$g(text_i) => [P_1^i, P_2^i, ..., P_k^i, P_{gov}^i]$$

Overlaps between parties are hence captured by similarly structure probability vector. Left-leaning legislators from right-wing parties will exhibit higher probabilities to belong to left-leaning parties than their right-leaning counterpart. To obtain ideological estimates, we finally reduce the k-th dimensional probability vector to one dimension using PCA (note that we exclude the government probability from the PCA). These estimates are at the speech-level, but can be aggregated at the speaker or party level.

Unlike classical supervised text classification, there is no need for train-test split here. On the contrary, we seek overfitting. Estimating partisan overlaps requires to identify any partisan pattern. Overfitting should help address the risk of confusing topic mixtures with topic preferences. Indeed, two parties or legislators should have similar position only if they talk about a given topic in the same way. Embedding is here especially important to capture not only the topic but also the position. Accordingly, overfitting the model should prevent semantic overlaps to be induced by the simple mention of similar topics.

3 Empirical Analysis

3.1 Data

To illustrate partisan semantic overlaps; we use data from Canada (1990-2019), Spain (1995-2019), Germany (1991-2017), New Zealand (1987-2019), France (2007-2017) and United Kingdom (1988-2019). Most datasets are from the ParlSpeech database (Rauh, De Wilde, and Schwalbach 2017). The Canadian dataset was obtained on lipad.ca (Beelen et al. 2017) and the French dataset was generously provided by the team of lafabriquedelaloi. Detailed information on source, period and number of speeches are presented in the following table.

Table 1: Data Overview

Country	Period	Source	N_Speeches
Canada	1990-2019	LiPad	594156
France	2007-2017	Frabrique De la Loi	615105
Germany	1991-2018	ParlSpeech	176055
New Zealand	1987-2019	ParlSpeech	536494
Spain	1996-2018	ParlSpeech	107167
UK	1988-2019	ParlSpeech	1591798

3.2 Face Validity: Overall Results

Figure 1 entails the main descriptive results. For each dataset, three different types of results are shown. The first column provides speaker-level distributions for each party over the whole period. The second column represents the first two dimensions of the PCA (Speaker Level). The third column represents the median partisan estimates over time. At first look, partisan results seem face valid. In all countries, the parties are ranked accurately. Social-democratic parties are systematically scaled on the left of conservative parties. IU in Spain is on the far left, as the German “Left” and the French “Communist”. Liberal centrist parties, such as the LibDem in the UK, the FDP in Germany and the UDI and Nouveau Centre in France. Green parties are also well located at the centre-left in all countries. The SNP is also surprisingly well located on the centre-left of the scale. The obvious and biggest problem certainly concerns the independentist party from Québec, Bloc Québécois (BQ), which is scaled on the far left. Left right is not very salient for this party, so that its estimates should be broadly spread around the centre of the scale. The bidimensional representation shows that speeches from the BQ cluster far away from other parties, which indicate very few overlaps between the BQ and other Canadian parties. The few existing overlaps certainly concern the NDP.

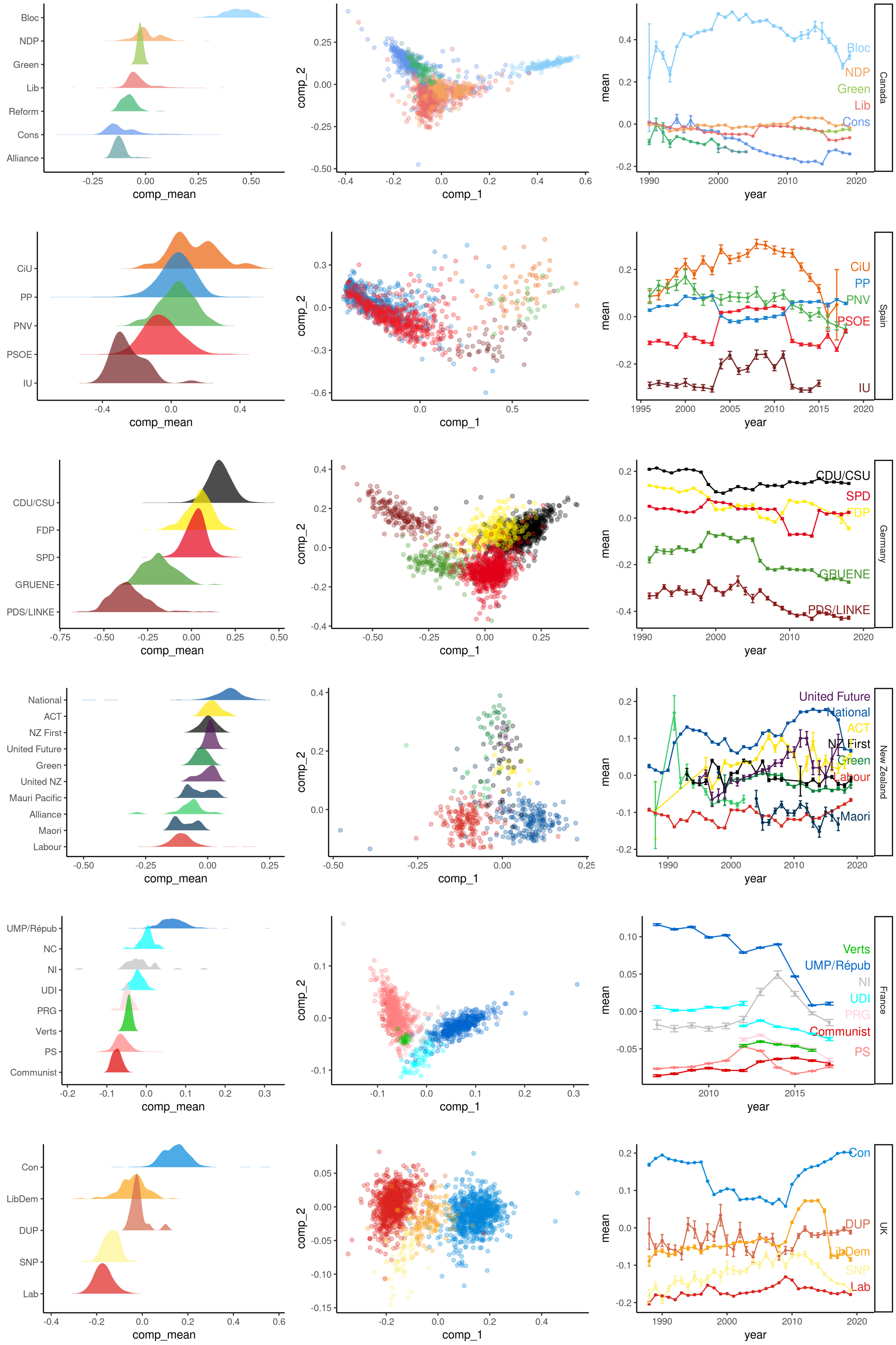

Figure 1: Main Descriptive results

3.3 Party-Level Validation: Comparative Manifesto Project

To confirm the face validity in a more rigorous way, we compare the speech estimates with the ideological scores proposed by the CMP.

To validate further those results, we look at the correlation between our estimates and the rile variable proposed by the CMP. As mentioned earlier, the rile variable is based on manifestos and provides an estimate of the ideological position of each party at each election. To match their data structure, speech estimates were aggregated at the partisan level over terms. The correlations between the two measures are presented in Figure 2. In most cases, we observe a strong correlation between the two series of estimates. We are here interested in the absolute value of the correlation regardless of its sign. The measures converge in Germany (.87), Canada (.69) and UK (.63). The multiplicity of parties makes the case of New Zealand complicated. In this context, the correlation coefficient of .49 is satisfying. The Spanish case proves to be challenging (.17). The correlation between our estimates and the CMP is really low. This could be explained by the high number of parties and the structural bi-dimensionality of the Spanish party system. Indeed, Spanish political competition is not only structured along the left-right dimension, but also along the centralization issue. While we would expect the semantic overlaps to represent this two-dimensional space, the first two components do not indicate to match these two dimensions. Further and more detailed investigation is needed here.

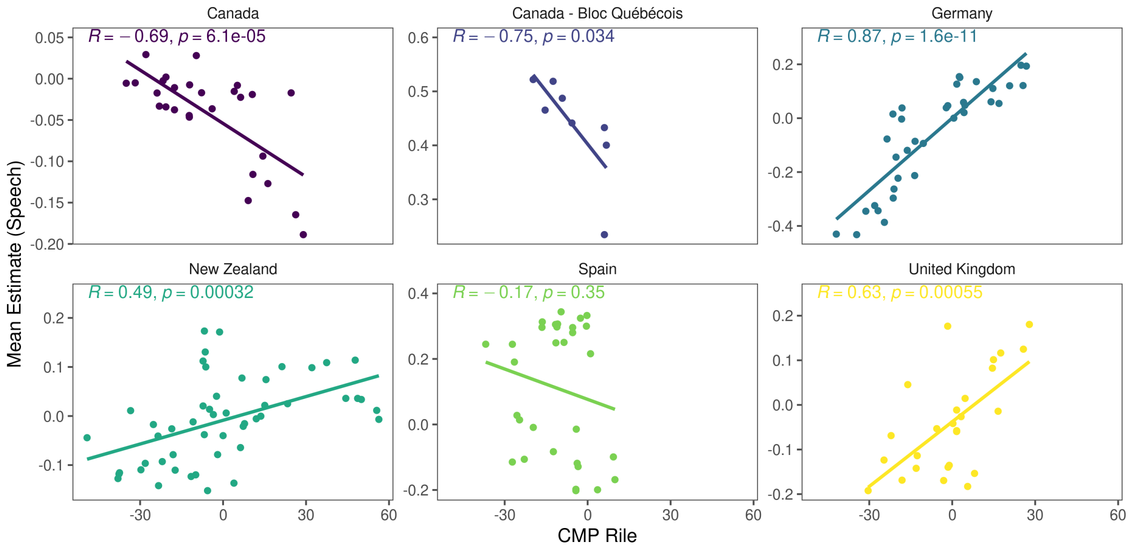

Figure 2: Party-Level Validation with CMP-Values

3.4 MP-Level Validation: German Expert-Survey on Individual Legislators

As mentioned earlier, individual measures of ideology need to be validated at the individual level. To do so, we take advantage of a separate study, which aims at measuring German MPs' ideological position using an expert survey. This survey asked young leaders of the German youth political parties to pairwisely compare 500 pairs of MPs given a left-right criteria. Using a Bradley-Terry model, these comparisons can be used to extract *ability* scores, which represents a latent factor, driving the probability of being labelled as *more left-leaning* than another MP. Although the survey scaled MPs from the current German parliament - on which we do not have speech data yet -, we used MPs that were already in office during the previous legislative period to match the two series of estimates. Results are presented in Figure 3 and 4. Overall, the results look very good. The measure correlate highly (.74). When disaggregating the correlation at the party-level, we find that the correlation is driven almost completely by accurate party estimates. Within parties, the two measures are almost independent from each other. These result seriously question the validity of the estimates, since within parties left-leaning MPs do not obtain left-leaning estimates.

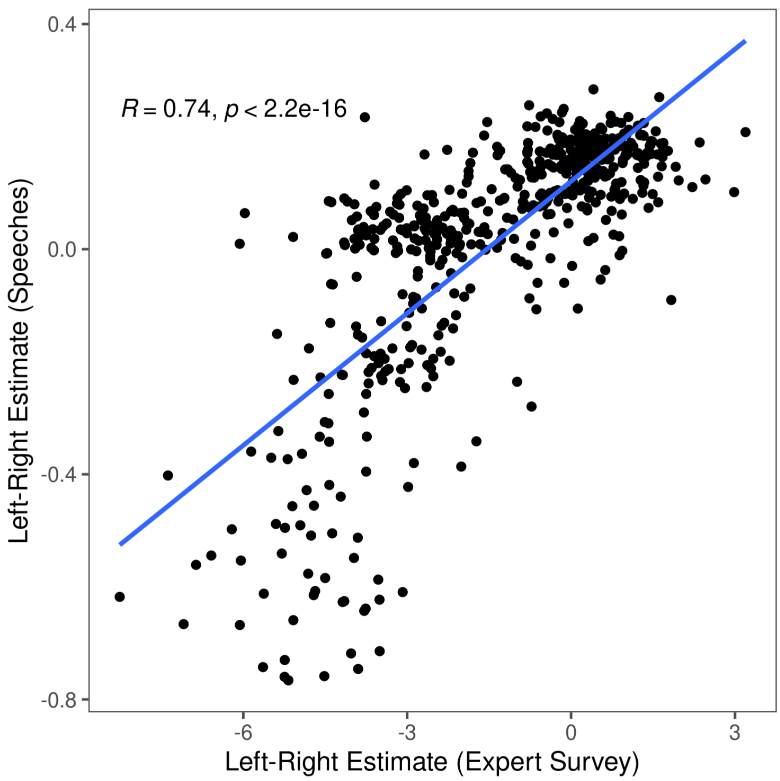

Figure 3: Individual Validation using German Expert Survey

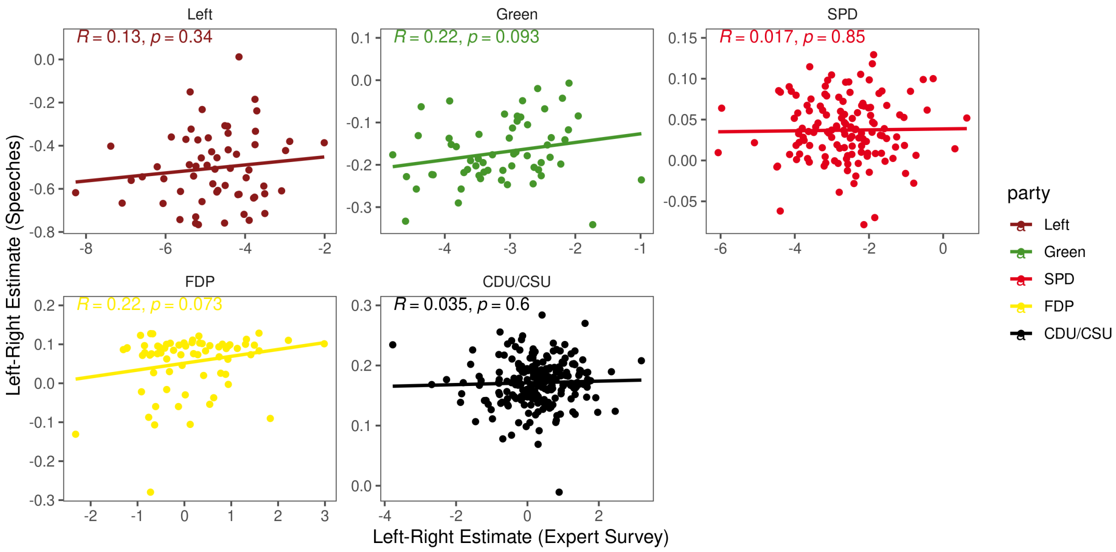

Figure 4: Individual Validation using German Expert Survey for each Party

4 Discussion

To conclude, we propose a new deductive behavioural measure of ideological position. In contrast to previous measures who modelled language, we rely on partisan semantic overlaps, measured as the probability of a given speech to belong to each party. We expect ideologically close members to hold similar speeches. To estimate these semantic overlaps, we use the prediction power of convolutional neural networks. We make them overfit so that the resulting probability amounts to indistinguishable language patterns. We validate the measure at the party and legislator levels. While parties are in most cases accurately estimated (high-correlation with the rile variable from the CMP), individual results are less convincing.

Those unsatisfying results can have several causes. First, the measure is perhaps conceptually wrong. Speeches, and thus semantic overlaps, are not necessarily driven by ideological position. Procedural speeches or technical ones are in essence for instance independent from ideology and cannot inform about their authors' position. Another way to explain the results is a lack of overfitting. Each model was trained with only two epochs and can overfit way more. We plan on training more epochs in the future, to see whether the individual results hold. Beyond these two explanations, topic might still be the issue. We expect the neural network to distinguish between different phrasing, but this requires enough data on any topic. Increasing overfitting and fine-tuning the model - using for instance sample weights, class weights, higher number of embedding dimensions - might help to overcome this issue. It might also help to add more data. The expert-survey run in Germany is currently run in UK and Canada. In these two countries, parliamentary speeches can be obtained over the whole last century. Focusing on UK and Canada might therefore provide a larger dataset and a way to validate the measure at the individual level. Finally, time-varying position might confuse the model. We chose to fit one model over all the period to maximize the amount of available information, but this might as well increase substantively the noise in the data. To sum up, these first results are encouraging but the measure needs better modelling and closer validation before it can be deployed to answer any substantial questions.

5 References

Beelen, Kaspar, Timothy Alberdingk Thijm, Christopher Cochrane, Kees Halvemaan, Graeme Hirst, Michael Kimmins, Sander Lijbrink, et al. 2017. "Digitization of the Canadian Parliamentary Debates." *Canadian Journal of Political Science/Revue Canadienne de Science Politique* 50 (3): 849–64.

Benoit, Kenneth, and Michael Laver. 2006. *Party Policy in Modern Democracies.* Routledge.

Carmines, Edward G, and Nicholas J D'Amico. 2015. "The New Look in Political Ideology Research." *Annual Review of Political Science* 18: 205–16.

Caughey, Devin, Tom O'Grady, and Christopher Warshaw. 2019. "Policy Ideology in European Mass Publics, 1981–2016." *American Political Science Review* 113 (3): 674–93.

Hooghe, Liesbet, Ryan Bakker, Anna Brigevich, Catherine De Vries, Erica Edwards, Gary Marks, Jan Rovny, Marco Steenbergen, and Milada Vachudova. 2010. "Reliability and Validity of the 2002 and 2006 Chapel Hill Expert Surveys on Party Positioning." *European Journal of Political Research* 49 (5): 687–703.

Kim, Heemin, and Richard C Fording. 1998. "Voter Ideology in Western Democracies, 1946–1989." *European Journal of Political Research* 33 (1): 73–97.

Lauderdale, Benjamin E, and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis*, 374–94.

Laver, Michael. 2014. "Measuring Policy Positions in Political Space." *Annual Review of Political Science* 17: 207–23.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–31.

Lowe, Will. 2008. "Understanding Wordscores." *Political Analysis*, 356–71.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *arXiv Preprint arXiv:1301.3781.*

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting.* Cambridge University Press.

Proksch, Sven-Oliver, and Jonathan B Slapin. 2009. "How to Avoid Pitfalls in Statistical Analysis of Political Texts: The Case of Germany." *German Politics* 18 (3): 323–44.

Rauh, Christian, Pieter De Wilde, and Jan Schwalbach. 2017. "The Parlspeech Data Set: Annotated Full-Text Vectors of 3.9 Million Plenary Speeches in the Key Legislative Chambers of Seven European States."

Rheault, Ludovic, and Christopher Cochrane. 2020. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* 28 (1): 112–33.

Sartori, Giovanni. 1970. "Concept Misformation in Comparative Politics." *The American Political Science Review* 64 (4): 1033–53.

Slapin, Jonathan B, and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22.

Volkens, Andrea, Pola Lehmann, Theres Matthiess, Nicolas Merz, Sven Regel, and Bernhard Wessels. 2017. "The Manifesto Data Collection: Manifesto Project Dataset (Version 2017a)." Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB).