PA

Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora

Ludovic Rheault[1] and Christopher Cochrane[2]

[1] Assistant Professor, Department of Political Science and Munk School of Global Affairs and Public Policy, University of Toronto, Canada. Email: ludovic.rheault@utoronto.ca

[2] Associate Professor, Department of Political Science, University of Toronto, Canada. Email: christopher.cochrane@utoronto.ca

Abstract

Word embeddings, the coefficients from neural network models predicting the use of words in context, have now become inescapable in applications involving natural language processing. Despite a few studies in political science, the potential of this methodology for the analysis of political texts has yet to be fully uncovered. This paper introduces models of word embeddings augmented with political metadata and trained on large-scale parliamentary corpora from Britain, Canada, and the United States. We fit these models with indicator variables of the party affiliation of members of parliament, which we refer to as party embeddings. We illustrate how these embeddings can be used to produce scaling estimates of ideological placement and other quantities of interest for political research. To validate the methodology, we assess our results against indicators from the Comparative Manifestos Project, surveys of experts, and measures based on roll-call votes. Our findings suggest that party embeddings are successful at capturing latent concepts such as ideology, and the approach provides researchers with an integrated framework for studying political language.

Keywords: word embeddings, parliamentary corpora, text as data, political ideology, natural language processing

1 Introduction

The representation of meaning is a fundamental objective in natural language processing. As a basic illustration, consider queries performed with a search engine. We ideally want computers to return documents that are relevant to the substantive meaning of a query, just like a human being would interpret it, rather than simply the records with an exact word match. To achieve such results, early methods such as latent semantic analysis relied on low-rank approximations of word frequencies to score the semantic similarity between texts and rank them by relevance (Deerwester *et al.* 1990; Manning, Raghavan, and Schütze 2009, chap. 18). The new state of the art in meaning representation is *word embeddings*, or word vectors, the parameter estimates of artificial neural networks designed to predict the occurrence of a word by the surrounding words in a text sequence. Consistent with the use theory of meaning (Wittgenstein 2009), these embeddings have been shown to capture semantic properties of language, revealed by an ability to solve analogies and identify synonyms (Mikolov, Sutskever, *et al.* 2013; Pennington, Socher, and Manning 2014). Despite a broad appeal across disciplines, the use of word embeddings to analyze political texts remains a new field of research.[1] The aim of this paper is to examine the reliability of the methodology for the detection of a latent concept such as political ideology. In particular,

1 Examples of recent applications in political research include Rheault *et al.* (2016), Glavaš, Nanni, and Ponzetto (2017), and Preoţiuc-Pietro *et al.* (2017).

we show that neural networks for word embeddings can be augmented with metadata available in parliamentary corpora. We illustrate the properties of this method using publicly available corpora from Britain, Canada, and the United States and assess its validity using external indicators.

The proposed methodology addresses at least three shortcomings associated with textual indicators of ideological placement currently available. First, as opposed to measures based on word frequencies, the estimates from our neural networks are trained to predict the use of language in context. Put another way, the method accounts for a party's usage of words, *given* the surrounding text. Second, our approach can easily accommodate control variables, factors that could otherwise confound the placement of parties or politicians. For example, we account for the government–opposition dynamics that have foiled ideology indicators applied to Westminster systems in the past and filter out their influence to achieve more accurate estimates of party placement. Third, the methodology allows us to map political actors and language in a common vector space. This means that we can situate actors of interest based on their proximity to political concepts. Using a single model of embeddings, researchers can rank political actors relative to these concepts using a variety of metrics for vector arithmetic. We demonstrate such implementations in our empirical section.

Our results suggest that word embeddings are a promising tool for expanding the possibilities of political research based on textual data. In particular, we find that scaling estimates of party placement derived from the embeddings for the metadata—which we call *party embeddings*—are strongly correlated with human-annotated and roll-call vote measures of left–right ideology. We compare the methodology with WordFish, which represents the most popular alternative for text-based scaling of political ideology. The two methods share similarities in that both can be estimated without the need for annotated documents. This comparison illustrates how embedding models are well suited to account for the evolution of language over time. In contrast to other text-based scaling methods, our methodology also allows to map political actors in a multi-dimensional space. Finally, we demonstrate how the methodology can help to advance substantive research in comparative politics by replicating the model in two Westminster-style democracies. For such countries in particular, the ability to include control variables proves useful to account for the effect of institutional characteristics.

We start by situating our methodology in the broader literature on textual analysis in the next section. We then introduce the methodology concretely in Section 3. Section 4 discusses preprocessing steps and software implementation. Next, an empirical section illustrates concrete applications of the methodology: we demonstrate strategies for the retrieval of ideological placement, validate the results against external benchmarks, and compare the properties of the method against WordFish. We also illustrate an application at the legislator level and briefly address the question of uncertainty estimates for prospective users. Finally, we discuss intended uses and raise some warnings regarding interpretation.

2 Relations to Previous Work

Two of the most popular approaches in political science for the extraction of ideology from texts are WordScores (Laver, Benoit, and Garry 2003) and WordFish (Slapin and Proksch 2008).[2] The first relies on a sample of labeled documents, for instance, party manifestos annotated by experts. The relative probabilities of word occurrences in the labeled documents serve to produce scores for each word, which can be viewed as indicators of their ideological load. Next, the scores can be applied to the words found in new documents to estimate their ideological placement. In fact, this approach can be compared to methods of supervised machine learning (Bishop 2006), where a computer is trained to predict the class of a labeled set of documents based on their

---

2 See also Lauderdale and Herzog (2016) for an extension of the method to legislative speeches and Kim, Londregan, and Ratkovic (2018) for an expanded model combining both text and votes.

observed features (e.g. their words). WordFish, on the other hand, relies on party annotations only. The methodology consists of fitting a regression model where word counts are projected onto party–year parameters, using an expectation maximization algorithm (Slapin and Proksch 2008). This approach avoids the reliance on expert annotations and amounts to estimating the specificity of word usage by party at different points in time.

Neither of these approaches, however, takes into account the role of words in context. Put another way, they ignore semantics. Although theoretically both WordScores and WordFish could be expanded to include *n*-grams (sequences of more than one word), this comes at an increased computational cost. There are so many different combinations of words in the English language that it rapidly becomes inefficient to count them. This problem has been addressed recently in Gentzkow, Shapiro, and Taddy (2016) and presented as a curse of dimensionality. Using a large number of words may be inefficient when tracking down ideological slants from textual data since a high feature–document ratio overstates the variance across the target corpora (Taddy 2013; Gentzkow, Shapiro, and Taddy 2016). But for a few exceptions (Sim *et al.* 2013; Iyyer *et al.* 2014), models of political language face a trade-off between ignoring the role of words in context and dealing with high-dimensional variables.[3] Instead of relying on word frequencies, word embedding models aim to capture and represent relations between words using co-occurrences, which sidesteps the high feature–document ratio problems while allowing researchers to move beyond counts of words taken in isolation.

In the context of studies based on parliamentary debates, an additional concern is the ability to account for other institutional elements such as the difference in tone between the government and the opposition. A dangerous shortcut would consist of attributing any observed differences between party speeches to ideology. In Westminster-style parliaments, the cabinet will use a different vocabulary than opposition parties due to the nature of these legislative functions. For instance, opposition parties will invoke ministerial positions frequently when addressing their counterparts. Hirst *et al.* (2014) show that machine-learning models used to classify texts by ideology have a tendency to be confounded with government–opposition language. As a result, temporal trends can also be obscured by procedural changes in the way government and opposition parties interact in parliament. A similar issue has been found to affect methods based on roll-call votes to infer ideology, where government–opposition dynamics can dominate the first dimension of the estimates (Spirling and McLean 2007; Hix and Noury 2016). Our proposed methodology can accommodate the inclusion of additional "control variables" to filter out the effect of these institutional factors.

We argue that neural network models of language represent the natural way for scholars to move forward when attempting to measure a concept such as political ideology. The patterns of ideas that constitute ideologies are not as straightforward as is often assumed (Freeden 1998; Cochrane 2015). Ideologies are emergent phenomena that are tangible but irreducible. Ideologies are tangible in that they genuinely structure political thinking, disagreement, and behavior; they are irreducible, however, in that no one actor or idea, or group of actors or ideas, constitutes the core from which an ideology emanates. Second-degree connections between words and phrases give rise to meaningful clusters that fade from view when we analyze ideas, actors, or subsets of these things in isolation from their broader context. People know ideology when they see it but struggle to say precisely what it is because they cannot resolve the irresolvable (Cochrane 2015). This property of ideology has bedeviled past analysis. Since neural network models are designed to capture complex interactions between the inputs—in our case, context words and indicator variables for political actors—they are well adapted for the study of concepts that should theoretically emerge from such interactions.

The neural network models used in this paper are derived from word embeddings (Mikolov, Chen, *et al.* 2013; Pennington, Socher, and Manning 2014; Levy, Goldberg, and Dagan 2015). Such

---

3 Other examples of text-based methods for the detection of ideology based on word and phrase occurrences include Gentzkow and Shapiro (2010), Diermeier *et al.* (2012), and Jensen *et al.* (2012).

models have gained a wide popularity in many disciplines, including, more recently, political science. For example, Preoţiuc-Pietro et al. (2017) use word embeddings to generate lists of topic-specific words automatically by exploiting the ability of embeddings to find semantic relations between words. Glavaš, Nanni, and Ponzetto (2017) utilize a word embedding model expanded to multiple languages as their main input to classify sentences from party manifestos by topic, which they evaluate against data from the Comparative Manifestos Project (CMP). Rheault et al. (2016) rely on word embeddings to automatically adapt sentiment lexicons to the domain of politics, which avoids problems associated with off-the-shelf sentiment dictionaries that often attribute an emotional valence to politically relevant words such as "health," "war," or "education." In this study, we expand on traditional word embedding models to include political variables (for an illustration using political texts, see Nay 2016). We examine how such models can be used to study the ideological leanings of political actors across legislatures.

3 Methodology

Models for word embeddings have been explored thoroughly in the literature, but we need to introduce them summarily to facilitate the exposition of our approach. This section also adopts a notation familiar to social science scholars. Our implementation uses shallow neural networks, that is, statistical models containing one layer of latent variables—or hidden nodes—between the input and output data.[4] The outcome variable w_t is the word occurring at position t in the corpus. The variable w_t is multinomial with V categories corresponding to the size of the vocabulary. The input variables in the model are the surrounding words appearing in a window Δ before and after the outcome word, which we denote as $\boldsymbol{w}_\Delta = (w_{t-\Delta}, \ldots, w_{t-1}, w_{t+1}, \ldots, w_{t+\Delta})$. The window is symmetrical to the left and to the right, which is the specification we use for this study, although nonsymmetrical windows are possible, for instance, if one wishes to give more consideration to the previous words in a sequence than to the following ones. Simply put, word embedding models consist of predicting w_t from $\boldsymbol{w}_\Delta$.

The neural network can be subdivided into two components. Let z_m represent a hidden node, with $m = \{1, \ldots, M\}$ and where M is the dimension of the hidden layer. Each node can be expressed as a function of the inputs:

$$z_m = f(\boldsymbol{w}'_\Delta \boldsymbol{\beta}_m). \tag{1}$$

In machine learning, f is called an activation function. In the case of word embedding models such as the one we rely upon, that function is simply the average value of $\boldsymbol{w}'_\Delta \boldsymbol{\beta}_m$ across all input words (see Mikolov, Chen, et al. 2013). Since each word in the vocabulary can be treated as an indicator variable, Equation (1) can be expressed equivalently as

$$z_m = \frac{1}{2\Delta} \sum_{w_v \in \boldsymbol{w}_\Delta} \beta_{v,m} \tag{2}$$

that is, a hidden node is the average of coefficients $\beta_{v,m}$ specific to a word w_v if that word is present in the context of the target word w_t. In turn, the vector of hidden nodes $\boldsymbol{z} = (z_1, \ldots, z_M)$ is the average of the M-dimensional vectors of coefficients $\boldsymbol{\beta}_v$, for all words v occurring in $\boldsymbol{w}_\Delta$:

$$\boldsymbol{z} = \frac{1}{2\Delta} \sum_{w_v \in \boldsymbol{w}_\Delta} \boldsymbol{\beta}_v. \tag{3}$$

Upon estimation, these vectors $\boldsymbol{\beta}_v$ are the *word embeddings* of interest.

---

4 For the purpose of our presentation, we follow the steps of the model that Mikolov, Chen, et al. (2013) call continuous bag-of-words (CBOW).
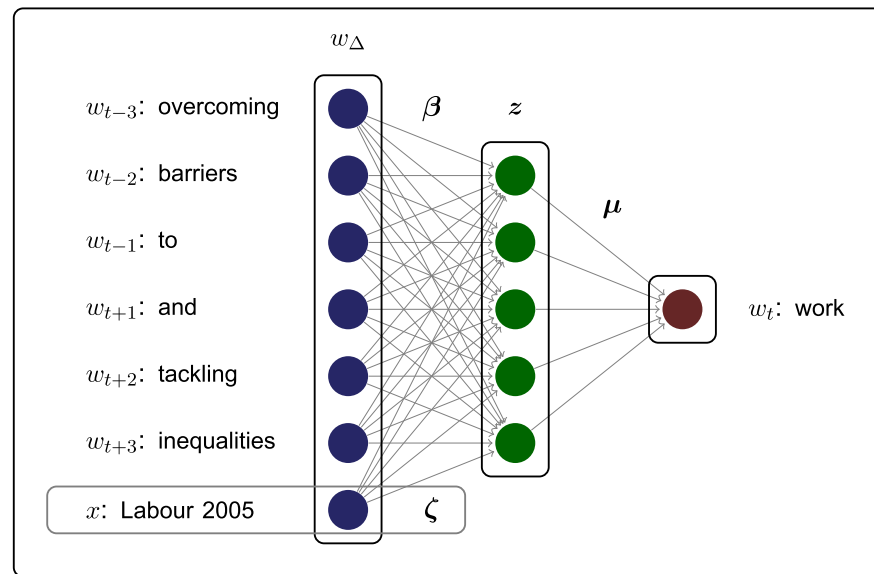
Figure 1. Example of model with word and party embeddings. The figure shows a schematic example of input and output data in a model with $M = 5$ and a window $\Delta = 3$. The model includes a variable indicating the party affiliation and parliament of the politician making the speech.

The remaining component of the model expresses the probability of the target word w_t as a function of the hidden nodes. Similar to the multinomial logit regression framework, commonly used to model vote choice, a latent variable representing a specific word i can be expressed as a linear function of the hidden nodes: $u_{it}^* = \alpha_i + \mathbf{z}'\boldsymbol{\mu}_i$. The probability $P(w_t = i)$ given the surrounding words corresponds to

$$P(w_t = i | \mathbf{w}_\Delta) = \frac{e^{\alpha_i + \mathbf{z}'\boldsymbol{\mu}_i}}{\sum_{v=1}^{V} e^{\alpha_v + \mathbf{z}'\boldsymbol{\mu}_v}}. \tag{4}$$

The full model can be written compactly using nested functions and dropping some indices for simplicity:

$$P(w_t | \mathbf{w}_\Delta) = g(\boldsymbol{\alpha}, \boldsymbol{\mu}, f(\mathbf{w}_\Delta' \boldsymbol{\beta})). \tag{5}$$

As can be seen with the visual depiction in Figure 1, the embeddings $\boldsymbol{\beta}$ link each input word to the hidden nodes.[5] The parameters of the model can be fitted by minimizing the cross-entropy using variants of stochastic gradient descent. We rely on negative sampling to fit the predicted probabilities in Equation (4) (see Mikolov, Sutskever, *et al.* 2013). In an influential study, Pennington, Socher, and Manning (2014) have shown that a corresponding model can be represented as a log-bilinear Poisson regression using the word–word co-occurrence matrix of a corpus as data. However, the implementation we use here facilitates the inclusion of metadata by preserving individual words as units of analysis.

The basic model introduced above can be expanded to include additional input variables, which is our main interest in this paper. A common implementation uses indicator variables for documents or segments of text of interest, in addition to the context words (Le and Mikolov 2014).[6] The approach was originally called paragraph vectors or document vectors. More generally,

---

5 In fact, Mikolov, Chen, *et al.* (2013) proposed two approaches: one in which the word embeddings are the link coefficients between input words and the hidden nodes (CBOW) and another where the outcome and the inputs are switched (called skip-gram)—in effect, predicting surrounding context from the word, rather than the reverse.

6 The type of model described here is called "distributed memory" in the original article (Le and Mikolov 2014).

other types of metadata can be entered in Equation (1) to account for properties of interest at the document level, which is the approach we adopt here. In our implementation, we focus primarily on indicator variables measuring the party affiliation of a member of parliament (MP) or a congressperson uttering a speech. The inner component of the expanded model can be represented as

$$z_m = f(\mathbf{w}'_\Delta \boldsymbol{\beta}_m + \mathbf{x}' \boldsymbol{\zeta}_m), \tag{6}$$

where $\mathbf{x}$ is a vector of metadata, and the rest of the specification is similar as before. In addition to party affiliation, it is straightforward to account for attributes with the potential to affect the use of language and confound party-specific estimates. We mentioned the government status earlier (cabinet vs. noncabinet positions or party in power vs. opposition). For Canada, a country where federal politics is characterized by persistent regional divides, a relevant variable would be the province of the MP. Just like words have their embeddings, each variable entered in $\mathbf{x}$ has a corresponding vector $\boldsymbol{\zeta}$ of dimension M.

Observe that the resulting vectors $\boldsymbol{\zeta}$ have commonalities with the WordFish estimator of party placement. In their WordFish model, Slapin and Proksch (2008) predict word counts with party–year indicator variables. The resulting parameters are interpreted as the ideological placement of parties. The model introduced in (6) achieves a similar goal. The key difference is that our model is estimated at the word level, while taking into account the context ($\mathbf{w}_\Delta$) in which a word occurs. The hidden layer serves an important purpose by capturing interactions between the metadata and these context words. Moreover, the dimension of the hidden layer will determine the size of what we refer to as *party embeddings* in what follows, that is, the estimated parameters for each party. Rather than a single point estimate, we fit a vector of dimension M. A benefit is that these party embeddings can be compared against the rest of the corpus vocabulary in a common vector space, as we illustrate below.

Specifically, our implementation uses party–parliament pairs as indicator variables for a number of reasons. First, fitting combinations of parties and time periods allows us to reproduce the nature of the WordFish model as closely as possible: each party within a given parliament or Congress has a specific embedding. This approach has relevant benefits by accounting for the possibility that the language and issues debated by each party may evolve from one parliament to the next. Parties are allowed to "move" over time in the vector space. We rely on parliaments/Congresses, rather than years, to facilitate external validity tests against roll-call vote measures and annotations based on party manifestos, which are published at the election preceding the beginning of each parliament. Of course, the possible specifications are virtually endless and may differ in future applications. But we believe that the models we present are consistent with existing practice and provide a useful ground for a detailed assessment.

4 Data and Software Implementation

Models of word embeddings have been shown to perform best when fitted on large corpora that are adapted to the domain of interest (Lai, Liu, and Xu 2016). For the purpose of this study, we rely on three publicly available collections of digitized parliamentary debates overlapping a century of history in the United States, Britain, and Canada. Replicating the results in three polities helps to demonstrate that the proposed methodology is general in application. The United States corpus is the version released by Gentzkow, Shapiro, and Taddy (2016). Our version of the British Hansard corpus is hosted on the Political Mashup website.[7] Finally, the Canadian Hansard corpus is described in Beelen *et al.* (2017) and released as linked and open data on www.lipad.ca. Each

---

7 See http://search.politicalmashup.nl/.

resource is enriched with a similar set of metadata about speakers, such as party affiliations and functions. The first section of the online appendix describes each resource in more detail.

We considered speeches made by the major parties in each corpus. The United States corpus ranges from 1873 to 2016 (43rd to 114th Congress). We present results for the House of Representatives and the Senate separately and restrict our attention to voting members affiliated with the Democratic and Republican parties. For the United Kingdom, the corpus covers the period from 1935 to 2014. We restrict our focus to the three major party labels: Labour, Liberal-Democrats, and Conservatives. For Canada, we use the entirety of the available corpus, which covers a period ranging between 1901 and 2018, from the 9th to the 42nd Parliament. The corpus represents over 3 million speeches after restricting our attention to five major parties (Conservatives, Liberals, New Democratic Party, Bloc Québécois, and Reform Party/Canadian Alliance). We removed speeches from the Speaker of the House of Commons in Britain and Canada, whose role is nonpartisan.

4.1 Steps for Implementation

We fit the embedding models on each corpora using custom scripts based on the library for Python of Řehůřek and Sojka (2010), which builds upon the implementation of document vectors proposed by Le and Mikolov (2014). This model itself extends the original C++ implementation of the word embeddings model proposed by Mikolov, Sutskever, et al. (2013). The library uses asynchronous stochastic gradient descent and relies on the method of negative sampling to fit the softmax function at the output of the neural networks. Our scripts are released openly, and the source code for the aforementioned library is also available publicly.[8] Fitting the models involves some decisions pertaining to text preprocessing and the choice of hyperparameters. We discuss each of these decision processes in turn.

4.1.1 Text Preprocessing

Preprocessing decisions when working with textual data may have nontrivial consequences (see Denny and Spirling 2018). Word embedding models are often implemented with little text preprocessing. One of the most popular implementations, however, that of Mikolov, Sutskever, et al. (2013), relies on subsampling—the random removal of frequent words from the context window during estimation. Subsampling essentially achieves a goal similar to stop word removal by limiting the influence of overly frequent tokens. This strategy was shown to improve the accuracy of word embeddings for tasks involving semantic relations. In our case, we want the models to learn information about terms that are substantively meaningful for politics. Simply put, the co-occurrence of the terms "reducing" and "taxes" is more meaningful to learn political language than that of the terms "the" and "taxes."

For this study, we preprocessed the text by removing digits and words with two letters or fewer, as well as a list of English stop words enriched to remove overly common procedural words such as "speaker" (or "chairwoman/chairman" in the United States), used in most speeches due to decorum. Even though our corpora comprise hundreds of millions of words, they are smaller than the corpora used in the original implementations of word embeddings. The removal of these common words ensures that we observe many instances of substantively relevant words used in the same context during the training stage. We tested models with and without stop words removed. Although removing procedural words only has a marginal effect on our methodology, we find that the removal of English stop words does improve the accuracy of our models for tasks such as ideology detection. We also limit the vocabulary to tokens with a minimum count of 50 occurrences. This avoids fitting embeddings for words with few training examples.

---

8   Our materials are available at Rheault and Cochrane (2019) and on a dedicated GitHub repository (https://github.com/lrheault/partyembed).

Finally, our models include not only words but also common phrases. We proceed with two passes of a collocation detection algorithm, which is applied to each corpus prior to estimation. Collocations (words used frequently together) are merged as single entities, which means that with two passes of the algorithm, we capture phrases of up to four words.[9] Since phrases longer than four words are very sparse, we stop after two runs of collocation detection on each corpus. Although not entirely necessary for the methodology, we find that the inclusion of phrases facilitates interpretation for political research, where multi-word entities are frequent and common expressions may have specific meanings (e.g. "civil rights"). The online appendix includes a table with the most frequent phrases for each corpus.

4.1.2 Fitting the Model

For the most part, we implemented our models using hyperparameters set at default values in the original algorithms proposed by Mikolov, Sutskever, *et al.* (2013) and Le and Mikolov (2014). We fit each model with a learning rate of 0.025 and five epochs—that is, five passes over all speeches in each corpus. Previous work using word embeddings often relies on hidden layers of size 100, 200, or 300. The main text reports models with hidden layers of 200 nodes, which we find to be reliable for applied research. The online appendix provides additional information on parameterization and its influence on the output. In essence, we find that modifications to these default hyperparameters do not provide substantial improvements to the results presented in this paper. Hence, choosing the values mentioned above appears to be a reasonable starting point and avoids overfitting the parameters to the characteristics of a specific corpus in ways that may not generalize over time.

The only departure from default hyperparameters is the choice of a window size. We rely on a window Δ of ± 20 words.[10] In contrast, implementations in the field of computer science often have window sizes of 5 or 10 words (see e.g. Levy, Goldberg, and Dagan 2015). This choice is based on our own examination of parliamentary corpora. We find that the discursive style of members of parliament is more verbose than the language typically found on the web. The chosen size roughly corresponds to the average length of a sentence in the British corpus (19.56 tokens per sentence on average) and the Canadian one (20.62 tokens per sentence). Moreover, the topics discussed in individual speeches tend to expand over several sentences. As a result, a window of 20 context words takes into account information from the previous and following sentences. We find this choice to be appropriate for our methodology. Our online appendix also reports sensitivity tests regarding the window size.

5   Empirical Illustrations

Upon estimation of the models, the embeddings can be used to compute several quantities of interest. In this section, we emphasize an approach to ideological scaling based on low-dimensional projections from the party embeddings.[11] We introduce tools facilitating interpretation and discuss validity assessments against three external sources of data on party ideology. Next, we compare the method against WordFish to emphasize a desirable property of word embedding models: the ability to account for changes in word usage over time. Finally, we illustrate how the methodology can be applied to individual legislators and discuss the question of uncertainty measures.

---

9 The algorithm used to detect phrases is based on the original implementation of word embeddings proposed in Mikolov, Sutskever, *et al.* (2013). Each pass combines pairs of words frequently used together as a single expression, for instance "united_kingdom." By applying a second pass, expressions of one word or two words can be merged, resulting in phrases of up to four words.

10 Note that during estimation, we weigh words according to their position in the context window, with the tth word weighted by a factor of $1/t$. At the beginning or end of a speech, the context window is truncated.

11 In the online appendix, we demonstrate how users can create projections on a customized space for the study of political ideology and compute other quantities of interest from the raw embeddings such as indicators of party polarization.

5.1 Party Embeddings

We start by assessing the models fitted with party-specific indicator variables. The objective is to project the M-dimensional party embeddings into a substantively meaningful vector space. These party embeddings can be visualized in two dimensions using standard dimensionality reduction techniques such as principal component analysis (PCA), which we rely upon in this section.

In plain terms, PCA finds the one-dimensional component that maximizes the variance across the vectors of party embeddings (see e.g. Hastie, Tibshirani, and Friedman 2009, chap. 14.5). The next component is calculated the same way by imposing a constraint of zero covariance with the first component. Additional components could be computed, but our analysis focuses on two-dimensional projections to simplify visualizations. If the speeches made by members of different parties are primarily characterized by ideological conflicts, as is normally assumed in unsupervised techniques for ideology scaling, we can reasonably expect the first component to describe the ideological placement of parties. The second component will capture the next most important source of semantic variation across parties in a legislature. To facilitate the interpretation of these components, we can use the model's word embeddings and examine the concepts most strongly associated with each dimension.

Starting with the US corpus, Figure 2(a) plots the party embeddings in a two-dimensional space for the House of Representatives. We label some data points using an abbreviation of the party name and the beginning year of a Congress; for instance, the embedding $\zeta_{\text{Dem 2011}}$ means the Democratic party in the Congress starting in 2011 (the 112th Congress). The only adjustment that may be relevant to perform is orienting the scale in a manner intuitive for interpretation, for instance, by multiplying the values of a component by -1 such that conservative parties appear on the right. Each model includes party–Congress indicator variables as well as separate dummy variables for Congresses, which account for temporal changes in the discourse. To facilitate visualization of the results, panels b and c in Figure 2 plot the party embeddings as time series, respectively, for the first and second component.

Our methodology captures ideological shifts that occurred during the 20th century. Although both major parties were originally close to the center of the first dimension, which we interpret as the left–right or liberal–conservative divide, they begin to drift apart around the New Deal era in the 1930s, a period usually associated with the fifth party system. The trend culminates with a period of marked polarization from the late 1990s to the most recent Congresses. The most spectacular shift is probably the one occurring on the second dimension, which we interpret as a South–North divide (we oriented the South to the bottom and the North to the top). The change reflects a well-documented realignment between Northern and Southern states that occurred since the New Deal and the civil rights eras (Shafer and Johnston 2009; Sundquist 2011). A similar trajectory is manifested using both the House and the Senate corpora (we report equivalent figures for the Senate in the online appendix). Although Republicans initially became associated with issues of Northern states, the opposite is true today—the two parties eventually switched sides completely. The recent era appears particularly polarized on both axes, which is consistent with a body of literature documenting party polarization. On the other hand, we do not find evidence of polarization on the principal component in the late 19th century, contrary to indicators based on vote data (Poole and Rosenthal 2007) but consistently with Gentzkow, Shapiro, and Taddy (2016).

5.1.1 Interpreting Axes

The models have desirable properties for interpreting the principal components in substantive terms. Since both words and party indicators are used as inputs in the same neural networks, we can project the embeddings associated with each word from the corpus vocabulary onto the principal components just estimated. Next, we can rank the words based on their distance from specific coordinates. For instance, the point (10, 0) in Figure 2 is located on the right end of the first
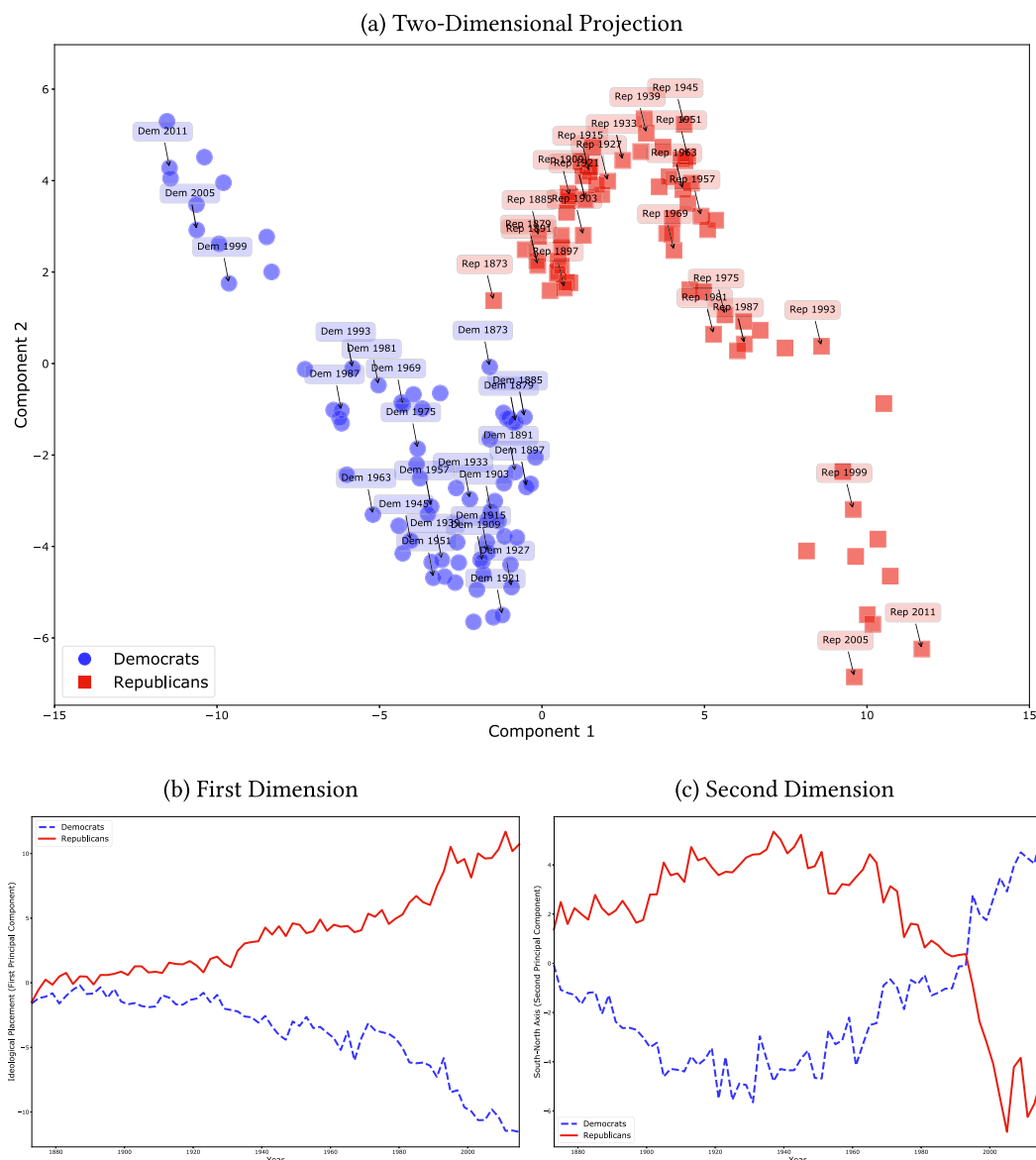
Figure 2. Party placement in the US House (1873–2016). The figure shows a two-dimensional projection of the two principal components of party embeddings for the US House of Representatives (a) and time-series plots for each of the two components separately in (b) and (c).

component. The words and phrases closest to that location can help researchers to interpret the meaning of that axis.[12]

To illustrate, Table 1 reports the 20 expressions with the shortest Euclidean distances to the four cardinal points of the two-dimensional space for the US House of Representatives.[13] We use the minimum and maximum values of the party projections on each axis to determine the cardinal points. Expressions located closest to the left end of the first component comprise "civil rights," "racism," "decent housing," and "poorest," indicating that these terms are semantically associated with this end of the ideological spectrum. These words refer to topics one would expect in the language of liberal (or left-wing) parties in the United States. Conversely, terms like "bureaucracy," "free enterprise," and "red tape" are associated with the right. As for the second

---

12 We also present an alternative tool for interpretation in the online appendix.
13 To help us focus on common expressions, we restrict our attention to the 50,000 most frequent words or phrases up to three words in the vocabulary.


Table 1. Interpreting PCA axes.

Component	Orientation	Words and phrases closest to edges of the PCA space
First	Positive (right)	Bureaucracy, bureaucracies, bureaucratic, Nebraska, regimentation, bureaucrats, Missouri River, Obamacare, centralized, redtape, Kansas, Hatch Act, charter schools, captive nations, free enterprise, Hoover Commission, Lenin, communist, feed grains, wheat
	Negative (left)	Congressional Black Caucus, wealthiest, decent housing, South African, slums, racism, poor elderly, African Americans, Latinos, African American, segregated, civil rights, gun violence, apartheid, African, poorest, joint resolution res, tax breaks, Brooklyn, richest
Second	Positive (North)	Detroit, Buffalo, Seattle, Minneapolis, Vermont, Duluth, Rochester, Lake Erie, debt gratitude, Lake Michigan, Erie, Cleveland, cleanup, retraining, Toledo, Maine, trade adjustment assistance, Oregon, Chicago, recycling
	Negative (South)	Brooks, usury, Sam Houston, parishes, Rome, cotton acreage, Poage, tobacco growers, Sam, burley tobacco, tyrannical, South Carolina, pink bollworm, southwest, Athens, fertilizer, military installations, North Carolina, Andrew Jackson, monopolies

dimension, the keywords refer to Southern and Northern locations or industries associated with each region, which supports our interpretation of that axis as a South–North divide.

We should emphasize that these lists may contain unexpected locutions, for instance, the "Missouri River" among the terms closest to the right edge of the first component. As argued earlier, political ideology cannot be reduced easily to any single idea or core. Attempting to summarize concepts such as liberal or conservative ideologies with individual words entails losing the context-dependent nature of semantics, which our model is designed to capture. For instance, some words contained in the lists of Table 1 may reflect idiosyncratic speaking styles of some Members of Congress on each side of the spectrum. Therefore, the interpretation ultimately involves the domain knowledge of the researcher to detect an overarching pattern. In this case, we believe the word associations provide relatively straightforward clues that facilitate a substantive interpretation of each axis. Since there are several different ways to explore relations between words and political actors in this model, an objective for future research would be to develop robust techniques for interpretation.

5.1.2 Replication in Parliamentary Democracies

Next, we illustrate that the methodology is generalizable across polities by replicating the same steps using the British and Canadian parliamentary corpora. To begin, Figure 3(a) reports a visualization of party embeddings for Britain. In addition to party and parliament indicators, the model includes a variable measuring whether an MP is a member of the cabinet or not. As can be seen, political parties are once again appropriately clustered together in the vector space: speeches made by members of the same group tend to resemble each other across parliaments. Moreover, the parties are correctly ordered relative to intuitive expectations about political ideology. Focusing on the first principal component (x-axis), the Labour party appears on one end of the spectrum, the Liberal-Democrats occupy the center, whereas the embeddings for Conservatives are clustered on the other side. In fact, without any intervention needed on our end, the model correctly captures well-accepted claims about ideological shifts within the British party system over time (see e.g. Clarke *et al.* 2004). For instance, the party embeddings for Conservatives during the Thatcher era (Cons 1979, 1983, and 1987) are ranked farther apart on the
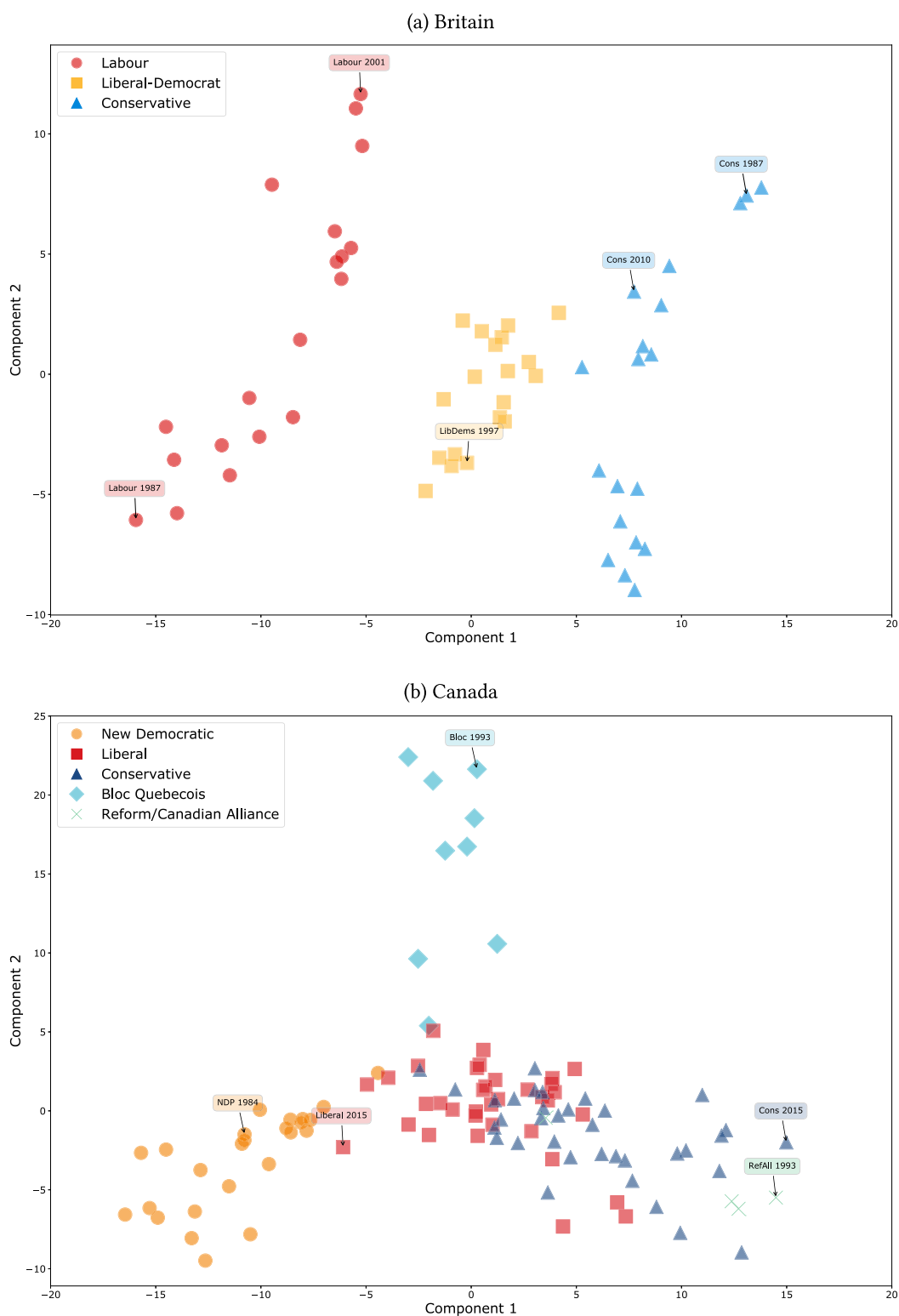
(a) Britain

(b) Canada

Figure 3. Party placement in Britain (1935–2014) and Canada (1901–2017). The figure shows the two principal components of party embeddings for the British and Canadian parliaments.

right end of the axis, whereas the Labour's shift toward the center at the time of the "New Labour" era (Labour 1997, 2001, and 2005), under the leadership of Tony Blair, is also apparent. The second component captures dynamics opposing the party in power and opposition, with parties forming a government appearing at the top of the y-axis.

Table 2. Accuracy of party placement against gold standards.

Gold standard	Metric	US House	US Senate	Canada	Britain
Voteview (1921–2016)	Correlation	0.918	0.919		
	Pairwise accuracy	85.66%	83.93%		
Expert surveys (1984–2002)	Correlation	0.982	0.981	0.869	0.910
	Pairwise accuracy	86.67%	100.00%	87.88%	83.33%
Rile (1945–2015)	Correlation	0.624	0.634	0.768	0.678
	Pairwise accuracy	72.61%	72.25%	77.03%	74.98%
Vanilla (1945–2015)	Correlation	0.736	0.736	0.731	0.755
	Pairwise accuracy	75.62%	75.09%	76.30%	78.06%
Legacy (1945–2015)	Correlation	0.898	0.907	0.855	0.876
	Pairwise accuracy	85.55%	85.37%	79.78%	82.67%

The first gold standard used for the US is the average DW-NOMINATE score (first dimension) from the Voteview project for House and Senate (Poole and Rosenthal 2007). Expert surveys are standardized measures of left–right party placement from three sources (Castles and Mair 1984; Huber and Inglehart 1995; Benoit and Laver 2006). The other three references are the Rile measure of party placement based on the 2017 version of the Comparative Manifestos Project (CMP) dataset (Budge and Laver 1992; Budge *et al.* 2001), the Vanilla measure of left–right placement (Gabel and Huber 2000), and the Legacy measure from Cochrane (2015). The pairwise accuracy metric counts the percentage of correct ideological orderings for all possible pairs of parties and parliaments.

Finally, Figure 3(b) reports the results for Canada.[14] Once more, the first principal component can be readily interpreted in terms of the left–right ideological placement. The Conservatives appear closer to the right, whereas the left-wing New Democratic Party (which is merged with its predecessor, the Co-operative Commonwealth Federation) is correctly located on the other end of the spectrum. The Reform/Canadian Alliance split from the Conservatives, generally viewed as the most conservative political party in the Canadian system (see Cochrane 2010), appears at the extreme right of the first dimension, consistent with substantive expectations. In the case of Canada, the second principal component can be easily interpreted as a division between parties reflecting their views of the federation. The secessionist party, the Bloc Québécois, appears clustered on one end of the *y*-axis, whereas federalist parties are grouped on the other side. Such a division also resurfaces in models based on vote data (see Godbout and Høyland 2013; Johnston 2017).

5.1.3 *Validity Assessments*

To assess the external validity of estimates derived from our models, we evaluate the predicted ideological placement against gold standards: ideology scores based on roll-call votes (for the US), surveys of experts, and measures based on the CMP data. For each gold standard, we report the Pearson correlation coefficient with the first principal component of our party embeddings. We also report the pairwise accuracy, that is, the percentage of pairs of party placements that are consistently ordered relative to the gold standard. Pairwise accuracy accounts for all possible comparisons, within parties and across parties. Table 2 presents the results.

Starting with the US corpus, we consider DW-NOMINATE estimates based on roll-call votes from the 67th to the 114th Congress and retrieved from the latest version of the Voteview project (Poole and Rosenthal 2007).[15] We use the first dimension of the aggregate Voteview indicator,

---

14 For that country, our model includes variables measuring whether the MP making the speech belongs to the cabinet or not, whether they belong to the party in power or the opposition, and their province of origin.

15 Note that for Westminster systems, vote-based estimates are less reliable indicators of ideology, and we did not consider them as a benchmark.

which measures the average placement of Congress members by party over time. Our ideological placement is strongly correlated with the Voteview scores ($r \approx 0.92$), and the pairwise accuracy, a more conservative metric, is around 85% for both the House and the Senate. These tests provide preliminary support to the conclusion that our model produces reliable estimates of ideological placement.

To further validate our methodology, we rely upon a second set of benchmarks based on human evaluations, namely expert surveys. Such surveys have been conducted sporadically in the discipline, asking country experts to locate national parties on a left–right scale. The average expert position is commonly used to examine party ideology at fixed points in time (see Benoit and Laver 2006). We retrieved expert surveys from three different sources covering the three countries under study in this paper (Castles and Mair 1984; Huber and Inglehart 1995; Benoit and Laver 2006). Two points on measurement should be emphasized. First, since expert surveys come from different sources, they may vary in construction and measurement scales. As a result, we standardize the scores before combining data points from different sources. We compute z-scores for each of the three surveys by subtracting the mean for all parties across the three countries and then dividing by the standard deviation. Second, we should point out that expert surveys provide us with only a few data points, in contrast to the other benchmarks reported in Table 2. In the United States, we retrieved expert party placements from three sources, which means six data points (i.e. Democrats and Republicans at three different points in time).

The third and fourth rows of Table 2 report the Pearson correlation coefficient and pairwise accuracy of our ideological placement—again, the first principal component of the party embeddings—evaluated against expert surveys. For both the US House and Senate, the two goodness-of-fit scores suggest that our methodology produces ideology scores consistent with the views of experts. The correlation coefficients are very high ($r \approx 0.98$), and the pairwise accuracy reaches 100% for the Senate. Despite the low number of comparison points, validating with a different source helps to give further credence to the interpretation of our unsupervised method of party placement. Experts surveys represent a more challenging test for the other two countries, which contain multiple parties, hence additional data points. The fit with expert surveys in Canada and Britain remains very strong, however, with correlation coefficients near or above 0.9.

Finally, we validate our ideological placement variables using data from the CMP (Budge and Laver 1992; Budge et al. 2001). The CMP is based on party manifestos and relies on human annotations to score the orientation of political parties on a number of policy items, following a normalized scheme. We test whether the three ideology indicators derived from the project's data are consistent with our estimated placement of the same party in the parliament that immediately follows. The Rile measure is the original left/right measure in the CMP. It is an additive index composed of 26 policy-related items, as described in Budge et al. (2001). The Rile metric, however, excludes important dimensions of ideology of the CMP, such as equality and the environment. The Vanilla measure proposed by Gabel and Huber (2000) uses all 56 items in the CMP and weights them according to their loadings on the first unrotated dimension of a factor analysis. Finally, the Legacy measure is a weighted index based on a network analysis of party positions and a model that assigns exponentially decaying weights to party positions in prior elections (Cochrane 2015).

Overall, the CMP-based indicators are consistent with our ideology scores, in particular when considering the Vanilla and Legacy measures. For the US, the correlation is positive but very modest when considering the more basic "right minus left" measure (Rile). Accuracy reaches $r \approx 0.9$, however, when considering the Legacy score.[16] Looking at the British case, party

---

16 Note that for the United States, the scores are only available for Presidential election years, whereas the corpus is divided into two-year Congresses. However, the CMP data are available from the election year 1920 in the US.
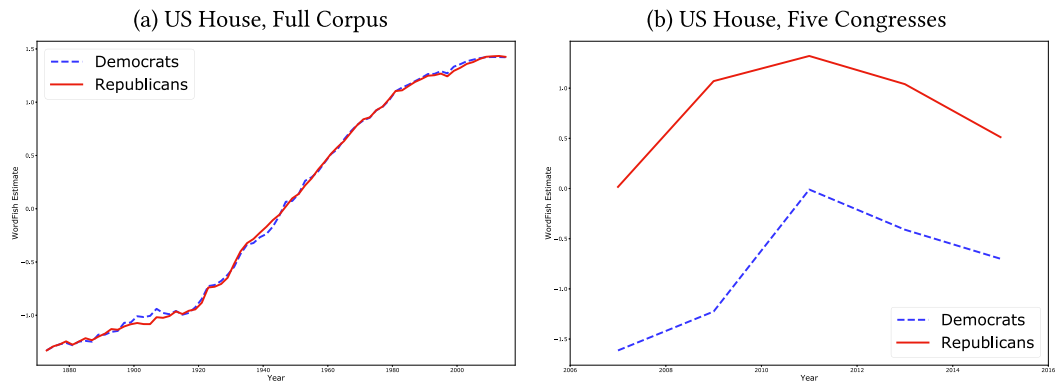
Figure 4. Comparison with WordFish estimates. The figure reports WordFish estimates for Democratic and Republican parties in the US House of Representatives, fitted on the full corpus (a) or using five Congresses from 2007 to 2015 (b).

placements appear positively correlated with the three external indicators, ranging from $r \approx 0.68$, when considering the Rile indicator, up to $r \approx 0.76$ and $r \approx 0.88$ using more robust ideology metrics based on the same data. As for pairwise accuracy, between 75% and 83% of comparisons against CMP-based measures are ordered consistently. Using the Canadian corpus, the fit with the CMP data is also strong across the three gold standards. Overall, these results provide strong evidence that our party embeddings are related to left–right ideology as coded by humans using party manifestos.

5.1.4 Comparison with WordFish

We now emphasize the properties of these embeddings using a comparison with estimates from the WordFish model.[17] The proposed methodology differs from WordFish in at least two important ways. Our model fits embeddings by predicting words from political variables and a window of context words. Each word in the corpus represents a distinct unit of analysis. In contrast, WordFish relies on word frequencies tabulated within groupings of interest. The data required to fit the WordFish estimator can be contained in a term–document matrix (for full expositions of this model, see Slapin and Proksch 2008; Proksch and Slapin 2010; Lowe and Benoit 2013). This matrix representation loses the original sequence in which the terms were used in the document, and the information loss is where the two models fundamentally differ. A second difference is that WordFish produces a single estimate of party placement, whereas the embeddings contain a flexible number of dimensions, which can be projected onto lower dimensional spaces, as illustrated above.

To illustrate these differences, we compute WordFish estimates on the entire US House corpus, restricting the vocabulary to the 50,000 most frequent terms.[18] We combined speeches by members of the same party into single "documents" for each Congress, following the approach used in Proksch and Slapin (2010). Figure 4(a) plots the estimated party positions for the full time period, which can be contrasted against our placement in Figure 2(b). When fitting the WordFish model on the full corpus, the estimator fails to capture the ideological distinctiveness of the parties. In fact, the dominant dimension appears to be the distinction between the language used in the late 19th century and that of more recent Congresses. Put another way, the estimates for both parties appear confounded by the change in language. Both parties are located next to each

---

17 Grimmer and Stewart (2013) have previously illustrated a limitation to WordFish when applied to individual legislators in Congress, a case for which the model does not perform well at capturing the expected ideological distance expected between members of opposite political parties.

18 We use the implementation available in the *quanteda* package in R, which is based on Lowe and Benoit (2013).

Table 3. Accuracy of party placement in the US House: WordFish and party embeddings.

Metric	1921–2016		2007–2016	
	WordFish	Embeddings	WordFish	Embeddings
Correlation	−0.027	0.918	0.829	0.999
Pairwise accuracy	44.36%	85.66%	75.56%	88.89%

Accuracy scores for WordFish estimates and party embeddings against the DW-NOMINATE scores (first dimension) from the Voteview project (Poole and Rosenthal 2007). The pairwise accuracy metric counts the percentage of correct ideological orderings for all possible pairs of parties and Congresses.

other over time, and the estimates bear no resemblance to benchmarks from either the Voteview project or the results based on our party embeddings.

The sharp contrast reflects a central property of the proposed methodology—and of word embeddings more generally—namely the ability to place equivalent expressions at proximity in a vector space.[19] Because the names and issues debated in early Congresses tend to be markedly different from those used in recent Congresses, the Democrats (or Republicans) of the past may seem disconnected from the Democrats (Republicans) of the present when considering word usage alone. In contrast, word usage across parties within a specific time period will appear very similar since both parties are debating the same topics. By accounting for the fact that different words can express the same meaning, a model based on word embeddings can account for continuity in a party's position even when word usage changes over time. For example, an expression related to an insurrection in the late 19th and early 20th century, the term "Moros" (referring to the Moro Rebellion in the Philippines) has a word embedding similar to the term "ISIL" (the Islamic State of Iraq and the Levant).[20] Even though ISIL was never mentioned in debates from the 19th century, the embedding model is trained to recognize that both terms share a related meaning.

When fitting the WordFish estimator on a shorter time span (the five most recent Congresses in the corpus), the model then places the two parties farther apart, and estimates become more consistent with the ideological divide expected between Democrats and Republicans (see Figure 4(b)). We calculate the accuracy of both implementations of the WordFish estimator against the Voteview benchmarks in Table 3. Although WordFish does not perform well for studying long periods of time, a shorter model produces estimates that are closer to those achieved with the first principal component of the proposed methodology. Both methods have benefits and limitations, and we should point out that an advantage of WordFish is that the model can be estimated on smaller corpora, as opposed to embedding models, which require more training data.

5.2 Legislator Embeddings

The methodology introduced in this paper can be adapted to the study of individual legislators, rather than parties. The specification is similar to that of the model used for political parties, with the difference that legislator–Congress indicator variables replace the party–Congress ones. We illustrate such an implementation by examining a model comprising embeddings fitted at the level of individual Senators in the US corpus. Since the number of speeches per legislator is lower than for a party as a whole, we compensate by increasing the number of epochs during training. The model discussed below is trained with 20 epochs.

Figure 5 reports the two principal components of the embeddings associated with individual Senators in the 114th Congress. The figure illustrates the expected division of Senators along party lines, although the clustering is nowhere as tight as that typically obtained using scaling methods

---

19 We are indebted to an anonymous reviewer for their helpful suggestions regarding this discussion.
20 The cosine similarity between the two terms' embeddings is 0.48.
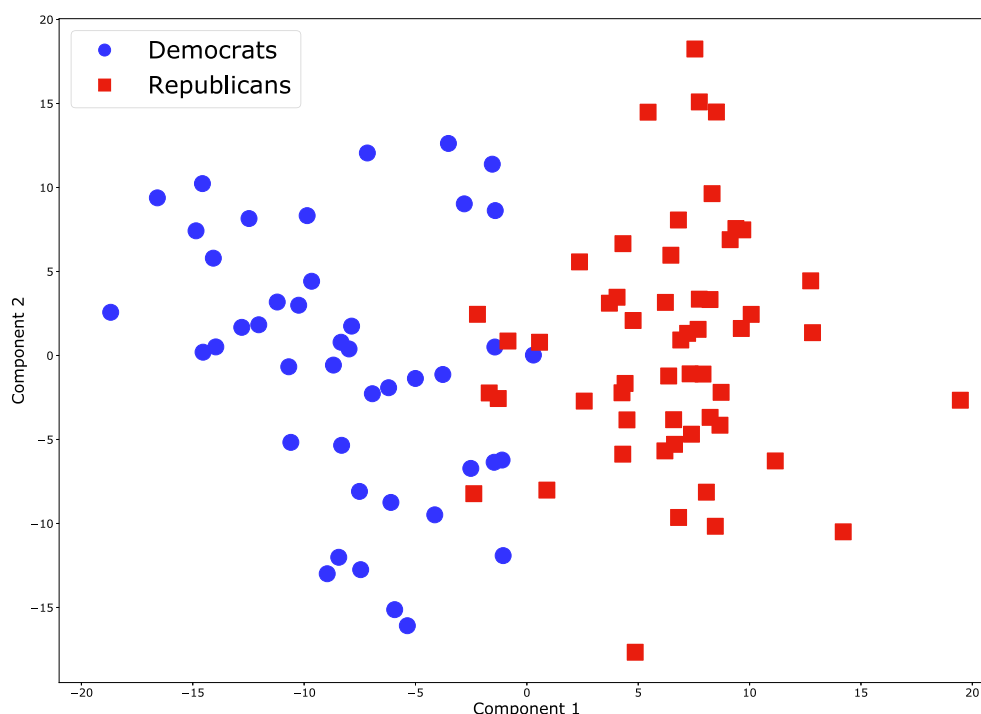
Figure 5. Ideological placement of Senators (114th Congress).

based on roll-call vote data. For instance, some Republican Senators appear mixed together with Democratic counterparts in the center of the spectrum. This can be explained in part by the fact that language reflects a much wider variance in positions than binary vote data, an issue that could be explored in future research. The centrist Republicans in Figure 5 include Robert Jones Portman, Lamar Alexander, and Susan Collins, who were generally ranked as moderates in external sources of ideology ratings discussed below.

To assess whether the low-dimensional projection captures political ideology, we proceed once more with a validation against external measures. By focusing on a recent Congress, we are able to use a variety of measures of ideological placement for US Senators: the Voteview scores for individual Senators, ratings from the American Conservative Union (ACU) and GovTrack.us ideology scores.[21] Table 4 reports Pearson correlation coefficients and pairwise accuracy scores using the first principal component of our model and each gold standard. Since our corpus is restricted to members affiliated with major party labels, this analysis excludes two independent Senators. For each gold standard considered, we obtain correlation coefficients of at least 0.85. The fit with the GovTrack ideology scores is particularly strong. Overall, these accuracy results support the conclusion that the first principal component extracted from our "Senator embeddings" is related to political ideology.

5.3 Uncertainty Estimates

A natural question regarding models based on word embeddings is whether they can be fitted with measures of uncertainty, such as standard errors and confidence intervals commonly reported in inference-based research. For instance, it would be convenient to express uncertainty around a prediction of a political party's placement. Unfortunately, deriving measures of uncertainty with neural network models, even with a relatively modest degree of complexity, remains an

---

21 The ACU ratings for 2016 were retrieved from http://acuratings.conservative.org/acu-federal-legislative-ratings/ on August 29, 2018. GovTrack report cards are also for 2016 and were retrieved the same day at https://www.govtrack.us/congress/members/report-cards/2016/senate/ideology.

Table 4. Accuracy of Senator ideological placement.

Gold standard	Pearson correlation	Pairwise accuracy (%)
DW-NOMINATE	0.876	80.81
Nokken–Poole	0.885	81.09
ACU 2016	0.852	75.38
ACU 2015	0.862	71.28
ACU Life	0.884	80.43
GovTrack	0.923	86.01

The DW-NOMINATE score is the main indicator from the Voteview data, and the Nokken–Poole is an alternative implementation based on Nokken and Poole (2004). The ACU ratings are for the year indicated or, alternatively, the life-long rating attributed to each Senator. The methodology used for the GovTrack ideology scores is described in the source website.

area of research under development (see Gentzkow, Kelly, and Taddy 2017). Bootstrap methods would be impractical due to the size of the corpora and the computing time needed to fit each model. Machine learning aims at prediction rather than inference, and the preferred validation methods are designed to assess model specification using predictive accuracy metrics, as we did above, rather than measures of uncertainty (Mullainathan and Spiess 2017). However, there are solutions available for producing measures of uncertainty for quantities of interest after accepting the raw embeddings as point estimates. For example, when computing cosine similarities using lists of words, researchers working with word embeddings have relied on techniques such as permutation t-tests (Caliskan, Bryson, and Narayanan 2017) or bootstrap estimators (Garg *et al.* 2018).

6 Intended Uses and Caveats

The proposed methodology can help to support research agendas in a number of subfields of political science. In comparative politics, the use of scaling estimates based on roll-call votes is often problematic when exported outside the United States, in particular for countries with high levels of party discipline or where the executive sits inside the parliament. The party embedding model introduced here affords comparative scholars with new opportunities to study party ideology, whenever a collection of digitized parliamentary corpora is available. Since the model can be adapted to the study of individual legislators, the technique also provides an additional resource for research on intra-party conflict, a central focus of recent research involving textual analysis of parliamentary speeches (see e.g. Proksch and Slapin 2015; Bäck and Debus 2016; Lauderdale and Herzog 2016; Schwarz, Traber, and Benoit 2017; Proksch *et al.* 2018). Next, the model is well adapted for research on political representation (Powell 2004; Bird 2010). In particular, the party embeddings can allow researchers to go beyond descriptive representation and examine how the language of political actors relates to issues and groups of interest. The results presented earlier suggest a connection between parties and minorities in the US Congress, and the method could be adapted for studies on the representation of other groups of interest across legislatures. Finally, the model can serve as a basis to develop additional methodological tools for textual analysis going beyond word counts, with the goal of studying political semantics.

Some words of caution are warranted, however, when interpreting text-based indicators of ideological placement. There are well-known limitations to scaling estimates based on roll-call votes, and those are often relevant for applications based on textual data. For instance, strategic agenda-setting decisions may limit opportunities to vote on some issues, and nonrandom abstentions during votes could obscure the true ideal point of a legislator (Clinton 2012). In a similar fashion, the distribution of issues in parliamentary corpora may be constrained by who controls the agenda during a given legislature, and some legislators could strategically refrain

from expressing their true opinions. These constraints on legislative debates may also vary across polities. For instance, the daily order of business in the Canadian House of Commons is dominated by the government, whereas other legislatures may offer extended opportunities to opposition parties. Researchers working with the proposed methodology should make sure to gain familiarity with the traits characterizing a legislature and the behavioral constraints that may affect their conclusions.

7 Conclusion

Methods based on word embeddings have the potential to offer a useful, integrated framework for studying political texts. By relying on artificial neural networks that take into account interactions between words and political variables, the methodology proposed in this paper is naturally suited for research focusing on latent properties of political discourse such as ideology. This approach has the benefit of going beyond word counts, which improves the ability to account for changes in word usage over time, in contrast to existing methods. The present paper relied on a custom implementation of word embeddings trained on parliamentary corpora and augmented to include input variables measuring party affiliations. As illustrated in our empirical section, the resulting "party embeddings" can be utilized to characterize the content of parliamentary speeches and place political actors in a vector space representing ideological dimensions. Our validity assessments suggest that low-dimensional projections from the model are generally consistent with human-annotated indicators of ideology, expert surveys, and, in the case of Congress, roll-call votes measures. Judging by the wide adoption of word embeddings in real-world applications based on natural language processing, we believe that the technique is bound to gain in popularity in the social sciences. In fact, our paper has probably scratched only the surface of the potential of word embeddings for political analysis.

Despite the promise of (augmented) word embeddings for political research, scholars should be wary of some limitations with the methodology. First, fitting the models requires a large corpus of text documents to fully benefit from their properties for semantic analysis. We find that corpora of digitized parliamentary debates appear reasonably large for this type of research, but smaller collections of texts may not achieve the same level of accuracy. Second, obtaining uncertainty estimates for the neural network models necessary to fit the embeddings would require further research. Although we have illustrated a variety of techniques for model validation, the methodology could eventually be implemented using frameworks like Bayesian neural networks and variational inference (see e.g. MacKay 1992; Tran *et al.* 2017). This would facilitate the estimation of confidence intervals, although at this stage the cost in terms of computational time remains prohibitive. Moreover, the quantities of interest to political scientists will likely be derivatives of the embeddings—for instance, a measure of similarity between words and parties—which implies compounding uncertainty estimates across the different steps involved. On the other hand, we have shown that the methodology can be used with a limited number of arbitrary decisions and produce reliable indicators of party placement, eliminating a source of variance associated with individual judgments.

Supplementary material

For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2019.26.

References

Bäck, H., and M. Debus. 2016. *Political Parties, Parliaments and Legislative Speechmaking*. New York: Palgrave Macmillan.

Beelen, K., T. A. Thijm, C. Cochrane, K. Halvemaan, G. Hirst, M. Kimmins, S. Lijbrink, M. Marx, N. Naderi, R. Polyanovsky, L. Rheault, and T. Whyte. 2017. "Digitization of the Canadian Parliamentary Debates." *Canadian Journal of Political Science* 50(3):849–864.

Benoit, K., and M. Laver. 2006. *Party Policy in Modern Democracies*. New York: Routledge.

Bird, K. 2010. "Patterns of Substantive Representation Among Visible Minority MPs: Evidence from Canada's House of Commons." In *The Political Representation of Immigrants and Minorities*, edited by K. Bird, T. Saalfeld, and A. M. Wüst. New York: Routledge.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Budge, I., H.-D. Klingemann, A. Volkens, J. Bara, and E. Tanenbaum. 2001. *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments (1945–1998)*. Oxford: Oxford University Press.

Budge, I., and M. J. Laver, eds. 1992. *Party Policy and Government Coalitions*. London: Palgrave Macmillan UK.

Caliskan, A., J. J. Bryson, and A. Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases." *Science* 356(6334):183–186.

Castles, F. G., and P. Mair. 1984. "Left–Right Political Scales: Some 'Expert' Judgments." *European Journal of Political Research* 12(1):73–88.

Clarke, H. D., D. Sanders, M. C. Stewart, and P. Whiteley. 2004. *Political Choice in Britain*. Oxford: Oxford University Press.

Clinton, J. D. 2012. "Using Roll Call Estimates to Test Models of Politics." *Annual Review of Political Science* 15:79–99.

Cochrane, C. 2010. "Left/Right Ideology and Canadian Politics." *Canadian Journal of Political Science* 45(3):583–605.

Cochrane, C. 2015. *Left and Right: The Small World of Political Ideas*. Montreal, Kingston: McGill-Queen's University Press.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. 1990. "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science* 41(6):391–407.

Denny, M. J., and A. Spirling. 2018. "Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26(2):168–189.

Diermeier, D., J.-F. Godbout, B. Yu, and S. Kaufmann. 2012. "Language and Ideology in Congress." *British Journal of Political Science* 42(1):31–55.

Freeden, M. 1998. *Ideology and Political Theory: A Conceptual Approach*. Oxford: Oxford University Press.

Gabel, M. J., and J. D. Huber. 2000. "Putting Parties in Their Place: Inferring Party Left–Right Ideological Positions from Party Manifestos Data." *American Journal of Political Science* 44(1):94–103.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences* 115(16):E3635–E3644.

Gentzkow, M., B. T. Kelly, and M. Taddy. 2017. "Text as Data." NBER Working Paper w23276.

Gentzkow, M., and J. M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1):35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy. 2016. "Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech." NBER Working Paper: 22423.

Glavaš, G., F. Nanni, and S. P. Ponzetto. 2017. "Cross-Lingual Classification of Topics in Political Texts." In *Proceedings of the 2017 ACL Workshop on Natural Language Processing and Computational Social Science*, 42–46. Association for Computational Linguistics.

Godbout, J.-F., and B. Høyland. 2013. "The Emergence of Parties in the Canadian House of Commons (1867–1908)." *Canadian Journal of Political Science* 46(4):773–797.

Grimmer, J., and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*. Berlin: Springer.

Hirst, G., Y. Riabinin, J. Graham, M. Boizot-Roche, and C. Morris. 2014. "Text to Ideology or Text to Party Status? In *From Text to Political Positions: Text Analysis across Disciplines*, edited by B. Kaal, I. Maks, and A. van Elfrinkhof, 93–116. Amsterdam: John Benjamins Publishing Company.

Hix, S., and A. Noury. 2016. "Government–Opposition or Left–Right? The Institutional Determinants of Voting in Legislatures." *Political Science Research and Methods* 4(2):249–273.

Huber, J., and R. Inglehart. 1995. "Expert Interpretations of Party Space and Party Locations in 42 Societies." *Party Politics* 1(1):73–111.

Iyyer, M., P. Enns, J. Boyd-Graber, and P. Resnik. 2014. "Political Ideology Detection Using Recursive Neural Networks." In *Proceedings of the 2014 Annual Meeting of the Association for Computational Linguistics*, 1113–1122. Association for Computational Linguistics.

Jensen, J., E. Kaplan, S. Naidu, and L. Wilse-Samson. 2012. "Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech." *Brookings Papers on Economic Activity* Fall:1–81.

Johnston, R. 2017. *The Canadian Party System: An Analytic History*. Vancouver: UBC Press.

Kim, I. S., J. Londregan, and M. Ratkovic. 2018. "Estimating Spatial Preferences from Votes and Text." *Political Analysis* 26(2):210–229.

Lai, S., K. Liu, J. Xu, and L. an Zhao. 2016. "How to Generate Good Word Embedding? *IEEE Intelligent Systems* 31(6):5–14.

Lauderdale, B. E., and A. Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24(3):374–394.

Laver, M., K. Benoit, and J. Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2):311–331.

Le, Q., and T. Mikolov. 2014. "Distributed Representations of Sentences and Documents." In *Proceedings of the 31st International Conference on Machine Learning*, edited by E. P. Xing and T. Jebara, II-1188–II-1196. PMLR.

Levy, O., Y. Goldberg, and I. Dagan. 2015. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics* 3:211–225.

Lowe, W., and K. Benoit. 2013. "Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark." *Political Analysis* 21(3):298–313.

MacKay, D. J. C. 1992. "A Practical Bayesian Framework for Backpropagation Networks." *Neural Computation* 4(3):448–472.

Manning, C. D., P. Raghavan, and H. Schütze. 2009. *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, 3111–3119. Neural Information Processing Systems Foundation.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." In *Proceedings of Workshop at ICLR*, 1–12. International Conference on Representation Learning.

Mullainathan, S., and J. Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2):87–106.

Nay, J. J. 2016. "Gov2Vec: Learning Distributed Representations of Institutions and Their Legal Text." In *Proceedings of the 2016 EMNLP Workshop on Natural Language Processing and Computational Social Science*, 49–54. Association for Computational Linguistics.

Nokken, T. P., and K. T. Poole. 2004. "Congressional Party Defection in American History." *Legislative Studies Quarterly* 29(4):545–568.

Pennington, J., R. Socher, and C. D. Manning. 2014. "Glove: Global Vectors for Word Representation." In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Association for Computational Linguistics.

Poole, K. T., and H. L. Rosenthal. 2007. *Ideology and Congress*. New York: Transaction Publishers.

Powell, G. B. 2004. "Political Representation in Comparative Politics." *Annual Review of Political Science* 7(1):273–296.

Preoţiuc-Pietro, D., Y. Liu, D. Hopkins, and L. Ungar. 2017. "Beyond Binary Labels: Political Ideology Prediction of Twitter Users." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 729–740. Association for Computational Linguistics.

Proksch, S.-O., and J. B. Slapin. 2010. "Position Taking in European Parliament Speeches." *British Journal of Political Science* 40(3):587–611.

Proksch, S.-O., and J. B. Slapin. 2015. *The Politics of Parliamentary Debate*. Cambridge: Cambridge University Press.

Proksch, S.-O., W. Lowe, J. Wäckerle, and S. Soroka. 2018. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* 0(0):1–35.

Řehůřek, R., and P. Sojka. 2010. "Software Framework for Topic Modelling with Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. European Language Resources Association.

Rheault, L., and C. Cochrane. 2019. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." https://doi.org/10.7910/DVN/K0OYQF, Harvard Dataverse.

Rheault, L., K. Beelen, C. Cochrane, and G. Hirst. 2016. "Measuring Emotion in Parliamentary Debates with Automated Textual Analysis." *PLoS ONE* 11(12): e0168843.

Schwarz, D., D. Traber, and K. Benoit. 2017. "Estimating Intra-Party Preferences: Comparing Speeches to Votes." *Political Science Research and Methods* 5(2):379–396.

Shafer, B. E., and R. Johnston. 2009. *The End of Southern Exceptionalism: Class, Race, and Partisan Change in the Postwar South*. Cambridge: Harvard University Press.

Sim, Y., B. D. L. Acree, J. H. Gross, and N. A. Smith. 2013. "Measuring Ideological Proportions in Political Speeches." In *Proceedings of the 2013 Conference on Empirical Methods of Natural Language Processing (EMNLP)*, 91–101. Association for Computational Linguistics.

Slapin, J. B., and S.-O. Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52(3):705–722.

Spirling, A., and I. McLean. 2007. "UK OC OK? Interpreting Optimal Classification Scores for the UK House of Commons." *Political Analysis* 15(1):85–96.

Sundquist, J. L. 2011. *Dynamics of the Party System*. Washington, DC: Brookings Institution Press.

Taddy, M. 2013. "Multinomial Inverse Regression for Text Analysis." *Journal of the American Statistical Association* 108(203):755–770.

Tran, D., M. D. Hoffman, R. A. Saurous, E. Brevdo, K. Murphy, and D. M. Blei. 2017. "Deep Probabilistic Programming." In *Proceedings of the 5th International Conference on Learning Representations*, 1–18.

Wittgenstein, L. 2009. *Philosophical Investigations*. West Sussex, UK: Blackwell.