Word Embeddings for the Estimation of Ideological Placement in Parliamentary Corpora

Ludovic Rheault† and Christopher Cochrane‡

Abstract

Word embeddings, the estimates from neural network models predicting the use of words in context, have now become inescapable in applications involving natural language processing and artificial intelligence. 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 a model of word embeddings augmented with political metadata and trained on large-scale parliamentary corpora from Britain and Canada. We illustrate how the methodology can be used to produce scaling estimates of party and politician ideological placement, and assess the hypothesis of party polarization in both countries. To validate the methodology, we compare the results against estimates from the Comparative Manifesto Project. We also discuss commonalities with other scaling estimators based on textual data and roll call votes. Overall, we find that the methodology produces meaningful indicators for studying substantive questions in political science.

Prepared for the 35th annual meeting of the Society for Political Methodology, Provo, Utah, July 18–21, 2018.

†Assistant Professor, Department of Political Science and Munk School of Global Affairs and Public Policy, University of Toronto.
‡Associate Professor, Department of Political Science, University of Toronto.
The representation of meaning is a fundamental objective in natural language processing. One of the most popular methodologies from the recent literature pursuing that goal 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, Chen, Corrado, and Dean 2013; Pennington, Socher, and Manning 2014). The methodology improves upon previous tools in the family of latent semantic analysis (Deerwester et al. 1990; Turney and Pantel 2010), and many applications involving artificial intelligence now rely on word embeddings to encode language. 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 word embeddings and their extensions for studying parliamentary corpora, which have become readily available to political scientists. We show that neural networks for word embeddings can be augmented with metadata available in such corpora, in particular indicator variables of party affiliations. We illustrate the properties of this method and assess its validity for the estimation of ideological placement, using two public collections of digitized parliamentary debates from Britain and Canada to fit the models.

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 derived from models 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. Among other things, this also allows us to examine differences in how parties talk about the same issues, rather than only the differences in the issues that parties talk about. Second, our approach can easily accommodate control variables, factors that could otherwise confound the placement of parties or politicians. In particular, we account for the government-opposition dynamics that have foiled ideology indicators 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. For example, 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 fitted embeddings—which we call party embeddings—are strongly correlated with human-annotated measures of left-right ideology. We distinguish between two

¹Examples of recent applications in political research include Rheault et al. (2016), Preotiuc-Pietro et al. (2017), and Glavaš, Nanni, and Ponizetto (2017).
approaches to estimate the placement of political actors on ideological dimensions. The first consists of using scaling techniques on the raw estimates, and does not involve the judgment of researchers. The second, which we call semi-supervised, relies on linear projections onto pre-defined scales by choosing a relevant set of political expressions. Our findings indicate that both approaches lead to similar levels of accuracy when assessed against external benchmarks of party placement, which means that scholars can safely prefer the option that requires no arbitrary decisions when analyzing the ideological space in a given polity. We also show that the methodology is particularly well suited to conduct analyses where the main interest is to assess the proximity between actors and their association with groups of concepts, for instance to measure party polarization and issue ownership.

After reviewing the commonalities between the methodology and other techniques for inferring ideology from textual data, we introduce the models and the corpora. Next, we illustrate implementations and assess the accuracy of the ideology estimates. We discuss some limitations and potential pitfalls of word embeddings for political research in the concluding discussion.

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). 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 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.

2See also Lauderdale and Herzog (2016) for an extension of the method to legislative speeches.
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). This problem is usually called overfitting in the machine learning literature. Problems associated with high-dimensionality often preclude the reliance on n-grams. Corpora such as parliamentary debates can easily contain millions of n-grams, many more than the number of documents (e.g. speeches) typically available for any given year.

Yet, capturing ideology in text ideally requires taking into account specific sequences of text. For instance, reducing taxes has a political meaning that is intuitively associated with right wing parties. The expression does not simply correspond to the juxtaposition of reducing and taxes; each word may be used individually in many other contexts, without the same implications in terms of ideology. Likewise, the word rights used in a discussion about the defense of minority rights and the defense of property rights may actually hint at opposite ideological stances. More often than not, textual cues relevant to identify ideology come under the form of multi-word expressions (Gentzkow and Shapiro 2010; Sim et al. 2013; Iyyer et al. 2014). Phrases themselves may be misleading without considering the broader context. For example, a politician may be “warning against the idea of reducing taxes for corporations”, suggestive of a left-wing ideological leaning despite the occurrence of a collocation such as reducing taxes. Inferring ideological positions from textual data requires more than a consideration of word occurrences, and even of n-gram occurrences. Capturing the semantics of a text—its meaning—is the real challenge, and one that we aim to address with the methods proposed in this paper.

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 party in power and the opposition. A dangerous shortcut would consist of attributing any observed differences between party speeches to ideology. In any given parliament, the party in power will use a different vocabulary than opposition parties due to the nature of these legislative functions. For instance, opposition parties in Westminster systems 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 a dimension of the estimates (Spirling and McLean 2007; Hix and Noury 2016).

A number of recent models have benefited from the increasing cross-fertilization coming from the field of machine learning. For instance, Iyyer et al. (2014) have used recursive neural networks to predict the ideological orientation of sentences based on a set of labeled phrases,
reaching a predictive accuracy of close to 70%. This is a promising method, but one that requires an extensive training sample of political texts annotated by humans. To identify ideological components of congressional debates, Diermeier et al. (2012) have used the features of support vector machines trained instead on a corpus labeled by party. A related method for the detection of ideology in text is derived from counts of partisan phrases (Gentzkow and Shapiro 2010; Jensen et al. 2012). Finally, a recent study also uses classifiers’ accuracy scores themselves as an indicator of polarization over time (Peterson and Spirling 2018). But for a few exceptions (Sim et al. 2013; Iyyer et al. 2014), these approaches again face a trade-off between ignoring the role of words in context and dealing with high-dimensional variables. Moreover, as with roll-call vote measures of party ideology, the distinction between partisan and institutional factors remains a source of concern in this type of applied research.

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 approach uses shallow neural networks, that is, statistical models containing one layer of latent variables—or hidden nodes—between the input and output data. 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 $\Delta$ before and after the outcome word, which we denote as $w_{\Delta} = (w_{t-\Delta}, ..., w_{t-1}, w_{t+1}, ..., w_{t+\Delta})$. The window is symmetrical to the left and to the right, which is the specification we use for this study, although non-symmetrical 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 $w_{\Delta}$.

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

$$z_m = f(w'_{\Delta} \beta_m)$$

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 $w'_{\Delta} \beta_m$ across all input words (see Mikolov, Chen, Corrado, and Dean 2013). Since each word in the vocabulary can be

---

3For the purpose of our presentation, we follow the steps of the model that Mikolov, Chen, Corrado, and Dean (2013) call continuous bag-of-words (CBOW).
treated as an indicator variable, Eq. (1) can be expressed equivalently as

$$z_m = \frac{1}{2\Delta} \sum_{w_v \in w_\Delta} \beta_{v,m}$$  \hspace{1cm} (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 $z = (z_1, ..., z_M)$ is the average of the $M$-dimensional vectors of coefficients $\beta_v$, for all words $v$ occurring in $w_\Delta$:

$$z = \frac{1}{2\Delta} \sum_{w_v \in w_\Delta} \beta_v$$  \hspace{1cm} (3)

Upon estimation, these vectors $\beta_v$ are the word embeddings of interest.

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 $u_{it}^* = \alpha_i + z' \mu_i$. The probability $P(w_t = i)$ given the surrounding words corresponds to:

$$P(w_t = i | w_\Delta) = \frac{e^{\alpha_i + z' \mu_i}}{\sum_{i=1}^{V} e^{\alpha_v + z' \mu_v}}$$  \hspace{1cm} (4)

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

$$P(w_t | w_\Delta) = g(\alpha, \mu, f(w_\Delta \beta))$$  \hspace{1cm} (5)

As can be seen with the visual depiction in Figure 1, the embeddings $\beta$ link each input word to the hidden nodes.\(^4\) The parameters of the model can be fitted by minimizing the cross-entropy using stochastic gradient descent. We rely on negative sampling to fit the predicted probabilities in Eq. (4) (see Mikolov, Sutskever, Chen, Corrado, and Dean 2013). In an influential study, Pennington, Socher, and Manning (2014) have shown that a corresponding model can be represented as a 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

\(^4\)In fact, Mikolov, Chen, Corrado, and Dean (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.
The approach was originally called paragraph vectors or document vectors. More generally, other types of metadata can be entered in Eq. 1 to account for properties of interest at the document level, which is the approach we adopt here (for an illustration using political texts, see Nay 2016). In our implementation, we focus on an indicator variable measuring the party affiliation of a member of parliament (MP) uttering a speech, in a given legislature, or alternatively a variable indexing the individual MPs themselves. The inner component of the expanded model can be represented as:

$$z_m = f(w'_\Delta \beta_m + x' \zeta_m)$$

where $x$ 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. In particular, the government status (party in power versus opposition, or cabinet versus non-cabinet positions) has been shown to dominate scaling methods for ideological placement based on vote data in countries outside the United States (Hix and Noury 2016). For Canada, a country where federal politics is characterized by persistent regional divides, we include the province of the MP as an additional variable in the model. Even in a country like the United States, accounting for such attributes can be relevant. For instance, a variable could be introduced to distinguish between the majority or minority status of the congressperson making a speech, and therefore eliminate a source of variation that

---

5The type of model described here is called “distributed memory” in the original article (Le and Mikolov 2014).
is distinct from partisan leanings. Just like words have their embeddings, each variable entered in \( z \) has a corresponding vector \( \zeta \) of dimension \( M \).

Observe that the resulting vectors \( \zeta \) are similar in purpose to the WordFish estimates of party placement. In their WordFish model, Slapin and Proksch (2008) predict counts of word occurrences 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 \( (w, \Delta) \) 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 \( M \). An obvious 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 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, rather than years, simply to facilitate external validity tests against annotations based on party manifestos, which are measured at the election preceding the beginning of each parliament. When considering MPs instead of parties, in a similar fashion we rely on pairs of MP-parliament as indicator variables of interest. 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.

Parliamentary Corpora

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 two public resources containing digitized parliamentary debates covering a century of history in Canada and Britain. The Canadian Hansard corpus is described in Beelen et al. (2017) and released as linked and open data on the Lipad website. The British Hansard corpus can also be accessed online via the Political Mashup website. Each resource is enriched with metadata about speakers and attributes such as party affiliations and functions. Table 1 reports summary statistics on each corpus.

For Canada, we use the entirety of the available corpus, which covers a period ranging be-
Table 1: Size of the Preprocessed Corpora

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Time-Range</th>
<th>Speeches</th>
<th>Sample Size (Words)</th>
<th>Vocabulary Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Britain</td>
<td>1936–2013</td>
<td>3.8M</td>
<td>253M</td>
<td>95,440</td>
</tr>
<tr>
<td>Canada</td>
<td>1901–2017</td>
<td>3.0M</td>
<td>188M</td>
<td>80,301</td>
</tr>
</tbody>
</table>

The corpus statistics are computed after performing the preprocessing steps described in this section.

between 1901 and 2017, 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 used substantive knowledge about the history of Canadian parties to group alternative affiliations under these five labels where appropriate. For instance, the Conservatives became the Progressive-Conservatives in 1942, and then the Conservative Party again after a merger with the Canadian Alliance in 2004. We treat these various denominations as a single entity. However, we account for the major split of the Conservatives leading to the creation of the Bloc and Reform Party (later Canadian Alliance) in 1993 since this represents a significant transformation of the party system. For simplicity, we excluded speeches from members of minor parties such as Labour, Social Credit and the Greens. Importantly, we removed all speeches from the Speaker of the House of Commons. Unlike their counterparts in the US Congress, the Speakers in Canada do not participate in substantive debates. As a result, speeches from the Speaker are uninformative in terms of ideology, and the party affiliation becomes irrelevant. For the United Kingdom, we focus on the period from 1936 to 2013. We restrict our focus to the three major party labels: Labour, Liberal-Democrats, and Conservatives. Once again, we removed the Speakers of the House, whose function is similar to that of the Canadian Speakers.

For both corpora, we fit models with hidden layers of 100, 200 and 300 nodes to compare their accuracy. Each model uses a window $\Delta$ of $\pm 15$ words and includes tokens with a minimum count of 50 occurrences. Our models include not only words, but also common phrases. We proceed with two passes to detect collocations (words used frequently together) and merge them as single entities, which means that we capture phrases of up to 4 words. This is especially useful for political research, where multi-word entities are frequent and common expressions may have specific meanings (e.g. “civil rights”). The models are fitted using custom scripts based on Řehůřek and Sojka (2010)’s implementation for Python and a default 0.01 learning rate. 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”,

---

*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 4 words. The algorithm used to detect phrases is based on the original implementation of word embeddings proposed in Mikolov, Sutskever, Chen, Corrado, and Dean (2013).*
which is used at the beginning of most speeches due to the decorum in both Houses. These steps are not entirely necessary, but we found that they bring a marginal improvement in terms of speed and they help simplify the models. We replicated the models discussed below without any preprocessing and the results are not substantially altered. All of our scripts, along with the estimated vectors, will be released publicly on a website upon publication.

Two Approaches to Ideological Scaling

We propose two different approaches to analyze political texts using the fitted party embeddings. The first approach consists of extracting the principal components from the party vectors and interpret them as estimates of ideological placement. As long as ideology is the main dimension along which political actors differ in terms of semantics, after accounting for control variables, this interpretation is plausible. We refer to this first approach as unsupervised since it does not require any reliance on prior knowledge on ideological leanings. In the second approach, we identify words or expressions that define the dimensions, and use them to create a customized vector space for the political actors. We call this second approach semi-supervised since the choice of words as anchor points to define the reduced vector space will have some influence on the findings.

Unsupervised

We start by illustrating the simplest approach. We consider the estimated party embeddings, that is, the coefficients from the neural network models associated with each party and parliament. 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 higher-dimensional vectors of interest (see e.g. Hastie, Tibshirani, and Friedman 2009, Ch. 14.5). The second component is calculated the same way, by imposing a constraint of zero covariance with the first component. 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, then 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.

Starting with the British corpus, Figure 2 depicts the party positions in a two-dimensional space using the first two principal components. We label each party embedding using an abbreviation of the party name and the beginning year of a parliament; for instance, the embedding $\zeta_{\text{Labour} \ 1997}$ means the Labour party in the parliament starting after the 1997 election—during Tony Blair’s first government—and is labeled as “Labour 1997”. The only adjustment that may be rel-
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). For Figure 2, no such adjustments were necessary. In addition to the party-parliament indicator variable, the fitted model includes a binary variable indicating whether the MP is a cabinet member or not, and separate dummy variables for parliaments. The figure reports estimates from a model with 200 dimensions ($M = 200$), which we have found to perform best.

Figure 2: Party Placement in a 2D Space using Principal Components (Britain)

As can be seen from Figure 2, political parties are appropriately clustered together in the vector space: speeches made by members of the same group tend to resemble each other across parliaments, and the party embeddings are close to each other in the vector space. Moreover, the parties are correctly ordered relative to intuitive expectations about political ideology when 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 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. One particular point of interest is the first parliament of 1974, during which the Labour party appears closer to the center (Labour 1974.1). This may be caused in part by the lower number of speeches available during that short-lived parliament, but the session also represents a special case in that it featured a failed attempt at creating a ruling coalition, with the Labour eventually forming a minority government. The next Parliament (1974.2) appears to be the turning point for increasing polarization between Labour and Conservatives, a discussion to which we return later on. The second dimension (y-axis) is more difficult to assess substantively, but we interpret it as non-linear discrepancies between distinct eras with specific issues of contention. For instance, the pre-war parliament of the 1930s (when a National Coalition Government was formed) is clustered far apart from the others, while parliaments during the 1990s are closest to the other end of the scale.

To assess the validity of estimates derived from our model, we take the first principal component as our indicator of the left-right ideological placement of the parties and compare the values to data from the Comparative Manifesto Project (CMP) for the United Kingdom (Budge and Laver 1992). The CMP data are also based on text documents, but are derived from the annotations of human coders following a normalized scheme. We test whether the left-right score produced using a manifesto is consistent with the estimated placement of the same party in the parliament that immediately follows. We also compare the accuracy of indicators from models with various dimensions of the hidden layer ($M$). Table 2 shows the results. Still focusing on the British case, the party placements based on the model with 200 dimensions are positively correlated with three ideology indicators derived from the CMP data, ranging from $\rho \approx 0.73$ when considering the “right minus left” measure ($rile$) from the 2017 CMP dataset, up to $\rho \approx 0.78$ and $\rho \approx 0.84$ using more robust ideology metrics based on the same data. We also report the percentage of correct pairwise orderings, that is, whether our model ranks any pair of parties-parliaments in the same direction as for the CMP-based indicators, both across parties within the same parliament, and within parties across all parliaments. In the case of the model just discussed, close to 80% of the pairwise comparisons are ordered consistently with the CMP-based measures. These results suggest that the principal component of our party embeddings is strongly related to left-right ideology as coded by humans based on the party manifestos.

Figure 3 reports the results for Canada, replicating the same methodology just discussed using the Canadian Hansard. For that country, our model includes variables measuring whether the MP making the speech belongs to the party in power or the opposition, whether they are in the cabinet or not, and their province. Once again, the major party labels appear clustered together across parliaments in the vector space. The first principal component can be readily interpreted in terms of left-right ideological placement. The Conservatives appear consistently on the right, whereas the left-wing New Democratic Party (NDP, which is merged with its pre-
Table 2: Accuracy of Party Placement against CMP Gold Standard

<table>
<thead>
<tr>
<th>Method</th>
<th>Hidden Layer Size</th>
<th>rile</th>
<th></th>
<th>vanilla</th>
<th></th>
<th>legacy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised (PCA)</td>
<td>100</td>
<td>0.6946</td>
<td>75.40%</td>
<td>0.7162</td>
<td>75.54%</td>
<td>0.8159</td>
<td>79.61%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7295</td>
<td>76.92%</td>
<td>0.7760</td>
<td>78.23%</td>
<td>0.8449</td>
<td>81.57%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7185</td>
<td>75.91%</td>
<td>0.7550</td>
<td>77.50%</td>
<td>0.7948</td>
<td>79.25%</td>
</tr>
<tr>
<td>Semi-Supervised</td>
<td>100</td>
<td>0.6472</td>
<td>73.15%</td>
<td>0.6591</td>
<td>73.15%</td>
<td>0.7478</td>
<td>75.33%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7638</td>
<td>78.88%</td>
<td>0.7507</td>
<td>76.12%</td>
<td>0.7133</td>
<td>74.67%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.6502</td>
<td>73.95%</td>
<td>0.6630</td>
<td>73.66%</td>
<td>0.7365</td>
<td>75.83%</td>
</tr>
<tr>
<td>Unsupervised (PCA)</td>
<td>100</td>
<td>0.7418</td>
<td>77.28%</td>
<td>0.6954</td>
<td>75.85%</td>
<td>0.8196</td>
<td>80.30%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7577</td>
<td>77.93%</td>
<td>0.7213</td>
<td>76.44%</td>
<td>0.8410</td>
<td>80.82%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7511</td>
<td>76.66%</td>
<td>0.7150</td>
<td>75.43%</td>
<td>0.8427</td>
<td>80.27%</td>
</tr>
<tr>
<td>Semi-Supervised</td>
<td>100</td>
<td>0.7177</td>
<td>76.14%</td>
<td>0.6654</td>
<td>74.07%</td>
<td>0.8069</td>
<td>79.16%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7252</td>
<td>76.24%</td>
<td>0.6554</td>
<td>73.32%</td>
<td>0.8128</td>
<td>79.78%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7465</td>
<td>77.38%</td>
<td>0.6788</td>
<td>74.33%</td>
<td>0.8209</td>
<td>79.94%</td>
</tr>
</tbody>
</table>

The gold standards used are the *rile* measure of party placement based on the 2017 version of the Comparative Manifesto 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 *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 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. The Legacy measure is a weighted index based on a network analysis of party positions and a model that assigns exponentially decaying weight to party positions in prior elections (Cochrane 2015). The pairwise accuracy metric counts the number of correct ideological orderings for all possible pairs of parties and parliaments. The semi-supervised approach relies on the expressions in Table 3.

As we did for Britain, we report an assessment of the estimated party placements for Canada against the CMP data in the lower panel of Table 2. The correlation coefficients and pairwise accuracies demonstrate once again a strong fit between our approach and indicators from the CMP. The model with 200 dimensions achieves a correlation coefficient of $\rho \approx 0.84$ when considering Cochrane (2015)’s CMP-based legacy estimates of party placement. For both countries, our tests suggest that $M$ between 200 and 300 dimensions is optimal to achieve results consistent with
human judgments. In the appendix, we report additional tests of accuracy of the models using benchmarks from the literature on word embeddings.

**Semi-Supervised**

The second approach for retrieving party positions consists of defining the substantive nature of axes before proceeding with a lower-dimensional projection. To implement this method, we choose expressions representative of opposite ideological stances on economic and social issues. If more than one term is chosen to anchor a position, we can take the centroids for each group of words and phrases by averaging their embeddings. Finally, axes are created by taking the difference between the right and left centroids, for each dimension of interest. We project party embeddings onto the customized space by taking dot products:

\[
\zeta \cdot \left( \frac{\sum_{i \in L_{Right}} \beta_i}{V_{Right}} - \frac{\sum_{i \in L_{Left}} \beta_i}{V_{Left}} \right)
\]
where \( L_{\text{Left}} \) is the chosen lexicon for words identifying the left-wing, and \( V_{\text{Left}} \) the size of that lexicon (and similarly for the Right).\(^7\)

Figure 4 illustrates such a linear projection of party embeddings in a two-dimensional space for Britain. The neural network model is the same as that used in the previous subsection. The social dimension (y-axis) uses the expressions “civil rights” and “traditional values” to represent left and right, respectively. For the economic dimension (x-axis), we use the expressions “workers” and “commerce” for left and right. Consistent with expectations, the figure suggests that the Labour party is the left-most party in the United Kingdom, on both the economic and social dimensions. The Conservatives, on the other hand, are both socially and economically on the right.

Figure 4: Party Placement in a 2D Space using Customized Ideological Axes (Britain)

To assess the semi-supervised approach against the CMP benchmarks, which are unidimensional, we select a list of five words or phrases epitomizing the priorities of the left and the right, and for which equivalent expressions exist in both countries (see Table 3). We also try to avoid period-specific expressions in favor of persistent concepts. An interesting goal in future research would be to consider the temporal evolution of right-wing and left-wing concepts, an endeavor

\(^7\)This approach expands on a standard visualization technique for the analysis of word embeddings; for instance, a similar implementation is included in Google’s TensorBoard tool.
We adapted expressions to account for idiosyncrasies in the language of each House. For instance, “debt relief” is a frequent collocation in Britain but is not in the vocabulary of the Canadian corpus, whereas “debt reduction” is the closest substitute in Canada and is absent from the British vocabulary. The phrase “gap rich poor” is originally “gap between the rich and the poor.” That we had to leave aside due to space constraints. Ideally, the choice of words should be guided by theory whenever the objective is to place parties along dimensions of specific interest to scholars. For the phrases in Table 3, we focus on core ideas characterizing the emphasis on class divide and inequality to represent the left, whereas the right is defined by a focus on industry and the competitiveness of the economy.

We assess the accuracy of the resulting ideological placement, again for each party and parliament, using the same sources as we did for the unsupervised approach. We report the same two metrics in Table 2, in the rows labeled as “Semi-Supervised”. Overall, the strong positive correlations with indicators of left-right placement from the CMP suggest that the second approach does indeed capture an axis of ideological divide that is substantively meaningful. The correlations and pairwise accuracies suggest that the semi-supervised method, however, does not produce results more accurate than the unsupervised model. We tested many combinations of words to verify that this conclusion is not merely dependent on the expressions selected to create the linear projection (the full list of terms considered appears in Table A3 in the Appendix). In general, the simpler, unsupervised approach produces a fit to the benchmarks close to the one achieved with the semi-supervised method, if not better. Although it is possible for researchers to select expressions that maximize the fit with external data points such as CMP indicators, by expanding the search grid in the vocabulary and testing large numbers of configurations, we conclude that the unsupervised method is reliable enough for applied research.

### Mapping Members of Parliament

Rather than estimating party positions, the model can be adapted to fit MP-specific embeddings by including MP indicator variables into the model. To illustrate, we focus on the Canadian sample, for which we are able to compare the estimates to those obtained using scaling methods on roll call vote data (Godbout and Høyland 2017). Figure 5 depicts the estimated placement of Canadian MPs during the 40th Parliament, again using principal component analysis. We use
a color code to reflect party affiliations. Once more, the first component can be viewed as a traditional left-right ordering, with Conservative MPs being clustered on one end of the x-axis. There are no clear clusters grouping Liberal and NDP members apart from each other; in fact, members of both parties are mixed together around the lower values of the first component. This result may be partially explained by the nature of that parliament, during which both opposition parties threatened to topple the government and form a coalition, but we find that party members are generally not as tightly clustered as are the party embeddings discussed earlier.

Figure 5: MP Placement in a 2D Space using Principal Components (Canada)

Placement of MP embeddings for the 40th Canadian Parliament (2008-2011), projected on two dimensions using principal component analysis.

To be sure, positioning individual MPs based on their speeches yields strikingly different results from those obtained using algorithms based on votes. The degree of internal cohesion of Canadian parties during parliamentary votes, as measured with the classical Rice index, usually hovers around 0.99—that is, close to perfect party discipline—as is the case in several democracies (see e.g. Depauw and Martin 2008; Godbout 2014). In other words, Canadian MPs generally abide by an unbending party discipline, and vote data provide limited information on distances between MPs within a party. Moreover, scaling techniques performed on vote data in Westminster
systems have been shown to capture government-opposition dynamics as the primary dimension (Spirling and McLean 2007; Hix and Noury 2016). In contrast, our approach has the potential to uncover a rich variety of preferences expressed within legislatures. Whereas MPs risk punishment for breaking the party line during official divisions, language provides them with much more leeway to express dissensions and nuances, which may occasionally bring them closer to members of opposite parties. In fact, future research could take advantage of the discrepancies between voting and discursive patterns to further our understanding of legislative behavior in Westminster systems.  

Despite the limitations associated with vote based indicators of ideology in countries like Canada, we find a positive correlation between the first principal component depicted in Figure 5 and a vote-based measure of MP placement. Table 4 reports the correlation coefficient and pairwise accuracy of our indicator against estimates for the first dimension of Poole (2000)’s Optimal Classification (OC) algorithm applied to voting data from the 40th Canadian Parliament. The correlation coefficients range between 0.73 and 0.76, for models fitted with various vector sizes. We note, however, that positive correlations can be explained by the adjustment of the polarity in the OC algorithm: since we can use a right-wing MP to orient the dimensions, Conservative MPs appear consistently on the right end of the first dimension, as is the case for the our placement estimates. When reproducing linear projections based on the expressions of Table 3 and the semi-supervised approach we described above, the results become even less consistent with the OC estimates. This supports the idea that roll call vote measures applied to the Canadian House of Commons do not reflect divisions along a left-right axis. Because of the limitations of vote scaling techniques in legislatures with strong government-opposition dynamics, alternative indicators of MP placement would be required as a gold standard to assess the methodology further.

**Examining Issue Ownership**

By estimating parameters for political actors and words in a common vector space, our models are well suited for studying the language used by each actor in parliament. To illustrate, we focus on political issues. There is a sizable literature emphasizing the question of issue ownership, that is, whether political parties have a particular competence for handling a specific policy area (Bélanger 2003; Bélanger and Meguid 2008; Green and Hobolt 2008; Seeberg 2017). So far, a large part of that literature relies on indicators of perceived party competence derived from voter surveys. Our methodology affords researchers with an opportunity to examine the issues prioritized by political parties inside legislative bodies. To illustrate, we compute cosine similarities between topic vectors and the MP embeddings for the Canadian House of Commons. We create

---

8For a recent example of approach combining both textual and vote data, see Kim, Londregan, and Ratkovic (2018).
Table 4: Accuracy of MP Placement against Vote-Based Optimal Classification Scores

<table>
<thead>
<tr>
<th>Method</th>
<th>Hidden Layer Dimension</th>
<th>Pearson Correlation</th>
<th>Pairwise Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised (PCA)</td>
<td>100</td>
<td>0.7648</td>
<td>74.50%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7398</td>
<td>73.92%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7300</td>
<td>73.86%</td>
</tr>
<tr>
<td>Semi-Supervised</td>
<td>100</td>
<td>0.5202</td>
<td>60.80%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.5084</td>
<td>60.24%</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.5205</td>
<td>60.29%</td>
</tr>
</tbody>
</table>

The reference used is the Optimal Classification of MPs based on roll call votes during the 40th Parliament, fitted using the R library oc. The pairwise accuracy metric counts the number of correct ideological orderings for all possible pairs of MPs.

these topic vectors from a keyword describing a topic of interest—for instance “environment”—and expand that keyword into a topic lexicon by searching for the 20 closest synonyms. The efficient retrieval of synonyms is a feature of word embedding models, which means that we can easily expand concepts of interests by looking for neighboring words in the vector space. As a next step, we average the word embeddings across the generated list of topic words. This average vector represents the location of the topic in the vector space. Finally, we can measure the similarity of speeches made by an MP and a political topic by taking the cosine similarity between the MP embedding and the topic vector.

We illustrate this application by examining the extent to which members of the left-wing party in Canada (the NDP) have embraced the issue of environment over time. This question touches a broader theoretical debate on the evolution of left-wing parties, and whether the issues that they prioritize have changed over time (see Kitschelt and Hellemans 1990). Figure 6 plots the average cosine similarity between MPs affiliated with the NDP and an “environment” topic vector, using the method described above. We also report a 95% confidence interval around the mean. Since MPs discuss a variety of topics during a parliament, the similarity between MP embeddings and a specific concept or topic will not typically be high. However, the metric captures meaningful variations across parliaments. In particular, the trend in Figure 6 suggests that the issue of environment has indeed become more strongly associated with the speeches of MPs from the NDP over time. The first change occurs early on, in the 1960s, during the first wave of the conservation movement. The proximity of MPs with the topic becomes even higher in recent parliaments, supporting the view that environmental policy has become an important part of the party’s agenda.
Assessing Party Polarization in Britain and Canada

Another benefit of having estimates of party placement in a vector space is the possibility of computing quantities of interest based on metrics for vector distances. An obvious application of such metrics is the measurement of the degree of party polarization in a legislature over time. A recurrent finding in the political science literature is the increasing ideological polarization of political parties (McCarty, Poole, and Rosenthal 2006; Layman, Carsey, and Horowitz 2006; Abramowitz and Saunders 2008; Dalton 2008; Lee 2015; Cochrane 2015). In Canada, signs of expanding levels of ideological diversity can be found in the party platforms at least since the 1980s (Cochrane 2015). For Britain, however, previous research suggests that parties have depolarized since then Thatcher era (Clarke et al. 2009; Adams, Green, and Milazzo 2012). This question matters for the study of politics, insofar as polarization challenges a central theory in the discipline, stating that parties have an incentive to converge toward the center of the voter distribution (Black 1948; Downs 1957; Schofield and Sened 2006). Below, we reassess the hypothesis of polarization in Britain and Canada, based on our estimates.

Several metrics can be used with the embeddings to measure the distance between the language of political actors. One of the simplest is the Euclidean distance $d_{ij}$ between two vectors.
ζ_i and ζ_j, which is obtained as:

\[ d_{ij} = \sqrt{\sum_{m=1}^{M} (\zeta_{im} - \zeta_{jm})^2} \]  

(7)

For instance, we can use the party embedding for the Conservatives (ζ_i), and measure its Euclidean distance with the corresponding vector for the Labour (ζ_j) in a given parliament. Other distance metrics have been popularized for the analysis of word embeddings, such as Word’s Mover Distance (WDM) (Kusner et al. 2015). WMD measures the shortest path required to transform the words of a first document into the words of another document. The metric could be utilized for a variety of analyses using our model’s word embeddings, for instance to compare specific speeches or groups of speeches made by political actors. For simplicity, we focus on Euclidean distance in what follows.

To study party polarization in the United Kingdom, we adopt a definition similar to Peterson and Spirling (2018). That is, we assess polarization as the distance between the ideological positions of the two parties having formed the government since 1936, Labour and Conservatives. Figure 7(a) plots the Euclidean distance between the two party embeddings over time. The pattern is consistent with the expectation of a depolarization, and reflects some of the findings introduced earlier in Figure 2. We observe that speeches in the House are most distinct in the period starting with the Parliament after the second general election in 1974. The gap between the two major parties’ ideological placement is emphasized clearly until the 1997 election that brought the Labour party back in power.

Turning to the Canadian case, a comparable definition of party polarization (the distance between the two parties having formed the government) involves evaluating the distance between the Liberal and Conservative party embeddings across parliaments. The situation is complicated by the fact that a breakup of the party system occurred at the 1993 election, and the Reform Party/Canadian Alliance replaced the Conservatives in the opposition during three parliaments. We thus consider the positions of the Reform-Alliance, rather than the Conservatives, during that period of time. Figure 7(b) depicts the trend in party polarization in Canada, during the same time-span we used for Britain. The data support the claim of an increasing polarization in recent decades. A first peak occurs in the 35th Parliament (1993), and the language of the two major parties appears to drift apart once again after the mid-2000s.

**Discussion**

This paper has set out to examine a promising methodological tool for textual analysis and assess its applicability to the measurement of party ideology. Our approach relies on a custom model
Polarization is measured using the Euclidean distance between the party embeddings of the Labour and Conservative parties for Britain. For Canada, it is the distance between the party embeddings of the Liberal and Conservative parties (using the Reform/Canadian Alliance from the 35th to the 37th Parliament). The thick lines are smooth splines of the raw Euclidean distances.

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 benefit from all the properties of the embeddings for semantic analysis. On the other hand, once a model is estimated, the embeddings can readily be shared and reused across applications. Moreover, advances in computing power make it simple to fit large models within a few hours, even on a standard laptop computer. Second, the computation of uncertainty measures for machine learning predictions, such as the ones used in this paper, remains an evolving field of research. The complexity of the models often makes standard approaches such as bootstrapping impractical. However, scholars making use of measures derived from machine learning predictions can address some of these concerns using methods such as measurement error models. Third, our approach did not model language evolution explicitly. Although we
relied on a historical corpus of digital speeches, hence accounting for the political vocabulary used across the entire time-range, a research problem of interest consists of determining the extent to which expressions of political conflict change across epochs. Based on our estimates, for instance, we find that Canadian parties were much closer semantically in the first half of the 20th century. It could be that the main parties back in the days were really more similar than they are today, which is supported by research on the Canadian party system (see e.g. Godbout and Høyland 2013; Johnston 2017). But the conclusion may also be affected by differences in language that are not yet fully understood. Finally, it is worth emphasizing that the model is only as good as the substantive theory that guides the research. Especially when customizing axes of contention, the decision to anchor dimensions around certain words and expressions can be sensitive. In this paper, we emphasized the usefulness of a totally unsupervised approach, which we find to perform well and does not require arbitrary decisions.
References


Appendix

Evaluating Word Embeddings Trained on Parliamentary Corpora

Table A1 reports accuracy results for the word embeddings of the models used in the main text (fitted with $M = 200$). We rely upon public benchmarks commonly used to evaluate the methodology. Word embeddings can solve analogies of the type “Ottawa is to Canada as Paris is to…” (France), by subtracting the difference between the two vectors of a known relationship from the query vector for the incomplete one (Mikolov, Chen, Corrado, and Dean 2013). Using a challenging test containing over 3,000 analogies to solve, we obtain satisfying accuracy scores compared to models trained on larger corpora; in particular, the British corpus achieves a 67% accuracy rate. In comparison, the state-of-the-art achieved by Pennington, Socher, and Manning (2014) with Global Vectors (GloVe) was 75%, using a corpus of 42 billion words. Note, however, that we accounted for the smaller size of our corpora by restricting the tests to analogies containing words among the 10,000 most frequent in our vocabularies, to ensure that the model has had a minimal training with the expressions involved. Overall, the results confirm that our models perform well at capturing semantics, despite the smaller sample size.

Table A1: Word Embedding Accuracy - Analogy Tests

<table>
<thead>
<tr>
<th>Category</th>
<th>Canadian Hansard</th>
<th>British Hansard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Correct/Subtotal</td>
</tr>
<tr>
<td>Capitals: Common Countries</td>
<td>75.0%</td>
<td>(15/20)</td>
</tr>
<tr>
<td>Capitals: World</td>
<td>42.9%</td>
<td>(6/14)</td>
</tr>
<tr>
<td>Currency</td>
<td>0.0%</td>
<td>(0/2)</td>
</tr>
<tr>
<td>Family Relationships</td>
<td>54.8%</td>
<td>(23/42)</td>
</tr>
<tr>
<td>Grammar 1: Adjective-to-adverb</td>
<td>36.0%</td>
<td>(151/420)</td>
</tr>
<tr>
<td>Grammar 2: Opposite</td>
<td>57.9%</td>
<td>(139/240)</td>
</tr>
<tr>
<td>Grammar 3: Comparative</td>
<td>78.7%</td>
<td>(472/600)</td>
</tr>
<tr>
<td>Grammar 4: Superlative</td>
<td>86.4%</td>
<td>(95/110)</td>
</tr>
<tr>
<td>Grammar 5: Present-participle</td>
<td>72.1%</td>
<td>(173/240)</td>
</tr>
<tr>
<td>Grammar 6: Nationality-adjective</td>
<td>82.0%</td>
<td>(146/178)</td>
</tr>
<tr>
<td>Grammar 7: Past-tense</td>
<td>55.6%</td>
<td>(420/756)</td>
</tr>
<tr>
<td>Grammar 8: Plural</td>
<td>71.7%</td>
<td>(172/240)</td>
</tr>
<tr>
<td>Grammar 9: Plural-verbs</td>
<td>45.2%</td>
<td>(95/210)</td>
</tr>
<tr>
<td>Total</td>
<td>62.1%</td>
<td>(1907/3072)</td>
</tr>
</tbody>
</table>

Analogy tests based on a benchmark list of word associations for word embeddings, using the models fitted with 200 dimensions. To account for the smaller sample sizes, the models are evaluated by restricting to the vocabulary of the 10,000 most frequently observed words. The test sheet can be downloaded here.
The second accuracy test reported in Table A2 is another common benchmark based on a list of word similarities evaluated by humans (Finkelstein et al. 2002). The correlation coefficients measure to which extent the cosine similarities between the two words in our models are associated with human-based similarity scores for the same word pairs. For both models, we achieve positive and statistically significant correlation coefficients, using either Pearson’s method or Spearman’s rank-order correlation. The test comprises 353 word pairs.

Table A2: Word Embedding Accuracy - Word Similarity Tests

<table>
<thead>
<tr>
<th></th>
<th>Canadian Hansard</th>
<th>British Hansard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>0.5612 2.26e-27</td>
<td>0.5273 6.09e-24</td>
</tr>
<tr>
<td>Spearman Correlation</td>
<td>0.5858 3.20e-30</td>
<td>0.5422 1.79e-25</td>
</tr>
</tbody>
</table>

Analogy tests based on a list of 353 human-evaluated word similarities from Finkelstein et al. (2002). The test sheet can be downloaded here.

Semi-Supervised Approach

Table A3 reports the lists of terms tested when creating the semi-supervised indicator of left-right ideological placement, as discussed in the main text.

Table A3: Words and Phrases Tested for Semi-Supervised Ideological Placement

| Left | affordable housing, child poverty, gap rich poor, guaranteed minimum income, inequality, low income, workers, minimum wage, protective tariff, redistribution wealth, safety net, social security, social services, trade unions, working classes, gay lesbian, racial minorities, assisted dying, greenhouse gas, civil rights |
| Right| creating jobs, free enterprise, individual liberty, liberalis[z]e trade, private enterprise, private sector, cutting taxes [lower taxes], debt relief [debt reduction], taxpayers money, commerce, privatis[z]ation, competitive, industry, productivity, fight terrorism, law enforcement, tougher sentences, traditional values, secure borders, illegal immigrants |

The square brackets indicate differences in spelling and in word usage between the British and the Canadian corpora.