Word embeddings on ideology and issues from Swedish parliamentarians' motions: a comparative approach

Citation for the original published paper (version of record):
Freden, A., Johansson, M., Saynova, D. (2026). Word embeddings on ideology and issues from Swedish parliamentarians' motions: a comparative approach. Journal of Elections, Public Opinion and Parties, 36(2): 273-294.
http://dx.doi.org/10.1080/17457289.2024.2433979

Journal of Elections, Public Opinion and Parties

ISSN: 1745-7289 (Print) 1745-7297 (Online) Journal homepage: www.tandfonline.com/journals/fbep20

Word embeddings on ideology and issues from Swedish parliamentarians' motions: a comparative approach

Annika Fredén, Moa Johansson & Denitsa Saynova

To cite this article: Annika Fredén, Moa Johansson & Denitsa Saynova (2026) Word embeddings on ideology and issues from Swedish parliamentarians' motions: a comparative approach, Journal of Elections, Public Opinion and Parties, 36:2, 273-294, DOI: 10.1080/17457289.2024.2433979

To link to this article: https://doi.org/10.1080/17457289.2024.2433979

Published online: 04 Dec 2024.

JOURNAL OF ELECTIONS, PUBLIC OPINION AND PARTIES
2026, VOL. 36, NO. 2, 273–294
https://doi.org/10.1080/17457289.2024.2433979

Word embeddings on ideology and issues from Swedish parliamentarians' motions: a comparative approach

Annika Fredén[a], Moa Johansson[b] and Denitsa Saynova[b]

[a]Department of Political Science, Lund University, Lund, Sweden; [b]Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden

ABSTRACT
Quantitative analysis of large-scale political text data in the form of word embeddings has great potential for systematising differences between political parties. We examine the differences between embeddings obtained from speakers from the two competitors for the PM position in Sweden (Social Democrats and Moderates) over a 30-year period. The goal is to compare how off-the-shelf general pre-trained models perform relative to pre-training on a smaller dataset from the same domain. In the analysis, we focus on two types of concepts: issues and ideological terms. We find that generally, the off-the-shelf pre-trained models lead to more reliable results and greater emphasis on ideological differences between the studied parties.

ARTICLE HISTORY Received 14 February 2023; Accepted 16 September 2024

KEYWORDS Parliaments; text as data; word embeddings; machine learning

1. Introduction

A recent trend in the social sciences is to use machine learning techniques to survey political attention and conflict (Osnabrügge, Elliot, and Morelli 2021a; Rheault and Cochrane 2020; Rodman 2020; Rodriguez and Spirling 2022). Word embedding is one of these techniques: starting from a random initialization, a neural network is trained to compute a numeric vector for each word, based on the context in which the word occurs and thus capturing its meaning. Use cases in digital social science include the study of semantic relationships between words to investigate changes in word meaning over time (see e.g. Hamilton, Leskovec, and Jurafsky (2016), and Tahmasebi (2018); Bonafilia et al. (2023), for examples, in

CONTACT Denitsa Saynova saynova@chalmers.se Chalmers tekniska högskola, Data- och informationsteknik, 412 96 Göteborg, Sweden

Swedish). In the domain of political science, embeddings have been used to, for example, measure emotion in debates (Rheault et al. 2016), and as part of an analysis of ideological placement (Rheault and Cochrane 2020). Computing reliable word embeddings requires a lot of data, why computer scientists typically start with a process called *pre-training*. The aim is to get a reasonable word representation, corresponding to “everyday use” which can then be further trained, or *fine-tuned*, on a specialized dataset, revealing finer distinctions in meaning. In this study, we put particular focus on investigating the effects of different strategies both for pre-training and for fine-tuning. Previous -related studies have touched upon the issue of pre-training quite briefly. For example, Rodriguez and Spirling (2022) suggest that “off-the-shelf” pre-trained embeddings in general work well for analysing and categorizing material from parliamentary debates. However, their investigation primarily focuses on the difference between using off-the-shelf pre-trained embeddings versus skipping pre-training and instead directly train on the dataset of interest. They do not consider using domain-specific data for additional pre-training. Moreover, their study materials come from parliamentary debates with relatively clear political signalling, with focus on US congress debates.

The advantage of using off-the-shelf pre-trained embeddings is that these can simply be downloaded and used immediately. However, the embeddings are likely not trained on the same domain and vocabulary as the downstream task. On the other hand, training our own embeddings on domain-specific texts is more time-consuming and requires extra data collection and computing resources for customized pre-training. The number of domain-specific texts is also by definition smaller than the number of general texts. The present study therefore compares two approaches to investigate the trade-off between pre-training dataset size and pre-training dataset vocabulary:

(1) Off-the-shelf embeddings pre-trained on a *larger dataset* with a vocabulary more representative of common everyday language,
(2) Embeddings pre-trained on a *smaller dataset*, but with a vocabulary more closely matching our downstream task, here parliamentary language.

Our first contribution is thus an investigation into the trade-offs between dataset size and domain-specificity. A third option, not considered here, is to skip pre-training and start training directly on task data. This will however require much more training data than we have available, or many iterations of bootstrapped training, as in for example Rodman (2020).

A second technical contribution of our study is a more thorough investigation of how to perform fine-tuning and how to measure the stability and reliability of the resulting word embeddings. Antoniak and Mimno (2018) found that for small- to medium-sized datasets, it is typically necessary to bootstrap fine-tuning to get a measure of stability of the embeddings, as they otherwise risk to vary greatly with small variations in the training data. This is clearly undesirable if we plan to compare the word embeddings between, for example, different parties, as we do. We also investigate how many passes (epochs) over the fine-tuning data we should make and find that for low-signalling data such as parliamentary motions, more epochs than is common improves the results as this reveals subtle party-specific differences between word embeddings.

Looking at a proportional representation system, Sweden, where the debate climate is usually less straight-forward than in, for example,the UK, we make a hard case for our study. In general, parties under proportional representation systems need to collaborate in coalitions to reach a majority in parliament (Bäck and Bergman 2016). Therefore, they might express messages and policies more subtly. Still, the parties in Sweden align themselves clearly along a political left-to-right-dimension, and candidates tend to be more party-oriented under proportional representation than in plurality systems (Fernandes, Goplerud, and Won 2019; Høyland and Søyland 2019). However, polarity is less pronounced than in plurality systems such as the US. Different language processing methods may take these patterns into account in different ways.

We have chosen to run the analyses on data from parliamentarians' motions, as they usually signal new policies that the party representatives want to present to the parliament. These types of materials have been rarely used in natural language processing analyses before and should therefore be a useful complement to existing language processing political science studies, that, independently of context, focus heavily on legislative debates (Fernandes, Goplerud, and Won 2019; Høyland and Søyland 2019; Rodriguez and Spirling 2022). The test in the present study is on materials with early signals from the political elite, which can have large consequences for implementation of reforms and policies at a later stage.

The focus is the two main rivals for the PM position in the studied period – the Social Democrats on the left, and the Moderates on the right – which thus represents a government-opposition relationship. Our approach aims to scrutinize differences and similarities in party representatives' meaning and word use concerning several topics that are important for voters. Therefore, the word embeddings perspective analysing words in their context – should be more suitable than for example a sentiment analysis perspective, that focuses on explicit conflict and can rely more easily on dictionary-based approaches (Proksch et al. 2019).

We find that the different pre-training procedures produce different results in the way the parties' vocabularies are represented as word embeddings. They differ in terms of stability and focus – i.e. which types of words are picked up from the different parties. In line with previous studies, we find that the externally pre-trained word embeddings, in general, produce results that are relevant to understand and analyse political text materials and distinguish party characteristics. They perform better at revealing words that are ideological markers for the focal parties, such as *welfare* and *security*. However, the model pre-trained on a smaller domain-specific dataset does occasionally identify some additional concepts that are related to trends in the political discourse, such as *polarization*.

2. Technical foundations

Before discussing how word embeddings can be applied to extract meaning from political text, we introduce some foundational concepts related to neural networks: how they are trained in general, and how word embedding models work more specifically.

2.1. Neural networks

Neural networks are a widely used family of machine learning models that can be *trained* to approximate a function mapping some input(s) to

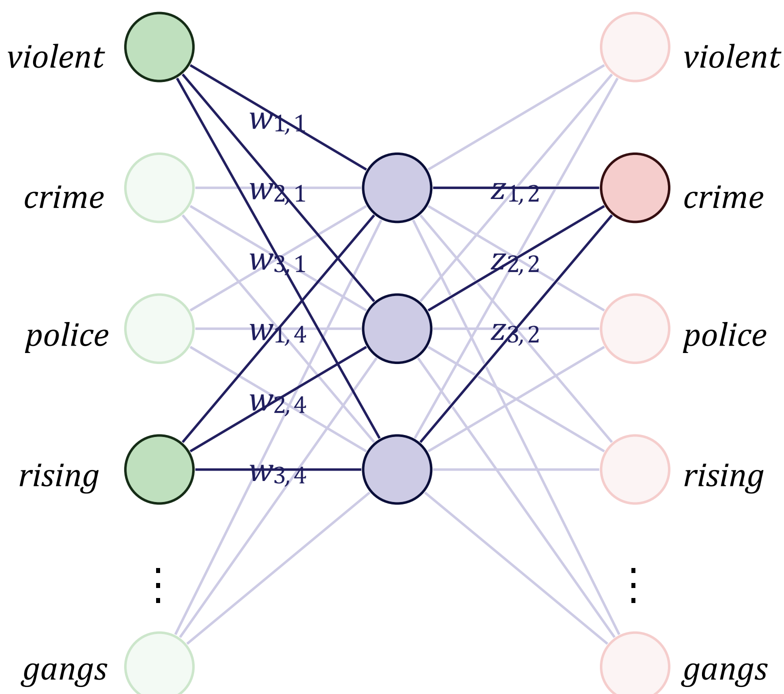

Figure 1. Word2vec: activation in the network when given the training example *violent crime rising*, with target *crime* and a context window of size 1.

outputs(s). A neural network consists of units called *neurones* organized in *layers* with connections with associated *weights* in between (see Figure 1 for an example). When training the neural network, the weights are incrementally updated to shift the output of the model closer to the correct value (the difference between correct and predicted answer is computed by a *loss function*).

There are several sources of randomness when training a neural network. For example, the weights are typically initialized randomly, so two runs of training may start from slightly different points. However, with a large enough training dataset, these effects become smaller, although one may expect ever so slightly different result from different runs. To remedy this, training can be bootstrapped over several iterations to average the results.

2.2. Embeddings and word2vec

Machine learning tools for text processing need to encode words as numerical values, called *word embeddings*, which should capture the meaning of a word as a numeric vector. In this work, we use one of the most common neural network-based methods, word2vec (Mikolov et al. 2013), specifically with its variant CBOW (Continuous Bag of Words). To learn the word embeddings, the neural network is trained to perform a proxy task: predicting a word from its context, where the context simply is the *n* closest words before and after the target word (the default value for word2vec is 5). As a small example, consider the phrase *violent crime rising* with target *crime* and context size 1. Figure 1 shows how this example will be passed through the network with the active connections in bold. Once the training is complete, the embeddings are extracted from the first layer of weights – the embedding for the word *violent* will be the vector $[w_{1,1}, w_{2,1}, w_{3,1}]$.

When a training example is passed through the network, the nodes corresponding to its context words will be active in the input layer. If this leads to the wrong word (not the target) becoming activated in the output layer, the associated weight will be updated to remedy this.

2.3. Training epochs

A *training epoch* is a pass over all the training examples once. Typically, more than one pass is needed for the model to learn: an early training example may trigger a weight update in a particular direction, while a later example may change them in a different direction. Therefore, multiple passes over the data help bring the weights to a state of compromise that can give a correct prediction on more and more examples. The optimal number of training epochs depends on the application and nature of the data and is often determined experimentally.

2.4. Pre-training and fine-tuning

Neural network models often require a lot of data, but it is not always necessary to start training from scratch, but rather from a *pre-trained* model. It is common that such models are trained once on some very large corpus, e.g. all Wikipedia articles, and then published for downstream tasks use. A second training stage, called *fine-tuning*, then trains the pre-trained model a bit more on task specific data (in our case text by the Social Democrats and Moderate parties respectively). The aim of pre-training is thus to learn general language patterns that are present in the type of text seen in the bigger corpus, ideally getting some representation of everyday use of the language. Fine-tuning on the other hand aims at slightly revising word embeddings to better fit language patterns in our domain-specific data. Note that the source of pre-training data matters, and generally, we want to have as much as possible of a vocabulary overlap between the pre-training data and the fine-tuning data for good results.

3. Materials and methods

Previous political science word embedding analysis tends to focus on only one or two key concepts, or a few randomly chosen concepts (Rodman 2020; Rodriguez and Spirling 2022). We develop their approaches and focus on two types of concepts: ideology more broadly and issues. Ideological terms are important since these contain political meaning, and issue words are important since they reveal political focus. In choosing concepts, we start from a latent semantic analysis of keywords that voters themselves mention when describing parties that they could vote for (Fredén and Sikström 2021). Seven concepts represented by ten keywords function as the starting point for the analysis. Three of them are concept words that are associated with ideology in a broader sense: *equality* [jämlikhet], *solidarity* [solidaritet], and *justice* [rättvisa], and four of these are words related to policy issues that are also potential political markers: (*drugs* [droger], *taxes* [skatt/er], *security* [säkerhet, trygghet][1] and *crime* [brott, brottslighet, kriminalitet]).

Some of these key concepts replicate the words that were used by Rodriguez and Spirling (2022) to detect the performance of different word embeddings in the US context (*justice*, *equality*, and *taxes*). The remaining concepts

[1]The concept of *security* is in the Swedish context related to two different dimensions: *state security*, referring to systems and programs [säkerhet]; and *social security*, referring to human feelings of safety and security [trygghet].

are representative for left- or right-leaning in the Swedish context, such as *crime* (issue-owned by the Moderates) and *solidarity* (issue-owned by the Social Democrats), and one issue, *drugs*, where party representatives' stances may be changeable (Bäck, Fredén, and Renström 2022).

3.1. Data

The starting point for our analyses is text corpora from private members' bills (motions) to the Swedish National Parliament (Riksdag).[2] We choose motions in the Swedish case because they reveal policy priorities and preferences at an early point in time, compared to the parliamentary debates, where an issue stance is usually set, compare (Osnabrügge, Hobolt, and Rodon 2021b). In these motions, parliamentarians, normally from the opposition, express their views on topics that are particularly relevant to them. Motions can thus be seen as representations of the politicians' discourses and party agendas during a specific period. Most importantly, we can distinguish *if* and *how* word meanings tend to converge or diverge between parliamentarians from the main parties in Sweden.

The studied period 1988–2020 contained the presence of six to eight parliamentary parties and the following shifts in office: Social Democrat 1988–1991, Moderates 1991–1994, Social Democrat 1994–2006, back to Moderate from 2006 until 2014. 2010 marks the entrance of the radical right-wing Sweden Democrats. The government shifted back to the Social Democrats in 2014 until 2020.

This work focuses on the full time span 1988–2020, to obtain a data set that is large enough and balanced in number of motions from each party. The distribution of number of documents and number of words by party can be seen in Table 1.

The frequency of the studied keywords in the training data can be seen in Table 2. The descriptive statistics indicate that the Social Democrats focus relatively more on solidarity and equality, whereas the number of motions concerning policy issues, in particular concerning taxes, is more frequent from the Moderates. In this analysis, minimal text processing has been done by lower casing the text and removing punctuation.

3.2. Pre-training

Two different approaches for pre-training are compared: a large off-the-shelf model and one that we pre-trained ourselves. The off-the-shelf pre-trained model comes from the Nordic Language Processing Library (NLPL) word

[2]http://data.riksdagen.se

Table 1. Number of motions for the data used in the analysis.

	Social Democrats	Moderates
Documents	31,360	30,056
Words	16,785,242	21,360,172

Table 2. Number of occurrences in the training data (1988–2020) for each of the investigated terms for each party.

Term	Translation	M	S
droger	drugs	724	508
skatt	taxes	6243	2190
säkerhet	state security	3658	2789
trygghet	social security	4033	2600
brott	crime	10466	3993
brottslighet	crime	2857	1686
kriminalitet	crime	841	509
jämlikhet	equality	184	485
solidaritet	solidarity	197	422
rättvisa	justice	812	1058

embedding repository.[3] It is trained on the Swedish CoNLL17 corpus and henceforth referred to as the NLPL model (Zeman et al. 2017). This model has a vocabulary size of 3,010,472 tokens.

The second model, which we pre-trained ourselves, is trained on additional data from the same domain as our task – in this case – texts from the Swedish Riksdag. This data is gathered from different written sources of governmental work: Interpellations (questions from MPs to ministers), Propositions (government proposals), and Parliamentary letters (notes on parliamentary decisions). After preprocessing, the final material batch contains 21,886 documents with 160,198,694 words and a vocabulary size of 1,042,282 unique words. Note that the number of documents is less than the motions data; however, these are longer documents, resulting in roughly 10 times more tokens.

In the NLPL model, we expect to see more general patterns of the Swedish language, while in the Riksdag model, we expect to find specific political and/or bureaucratic jargon and more politically charged terms. Still, the NLPL model is considerably larger. We want to investigate whether these differences will affect the resulting word vectors.

3.3. Related work

Rodman (2020) explores the development of the ideologically charged term *equality* in historical newspaper texts. Several sets of embeddings are trained,

[3]http://vectors.nlpl.eu/repository/20/69.zip

for different timeperiods using bootstrapping, to compensate for the relatively small dataset. Unlike our work, no pre-training is considered, as pre-trained models are typically available only for modern-day language. In contrast to Rodman, who focuses on one term, our work covers several politically relevant terms.

Rodriguez and Spirling (2022) compare different settings for computing word embeddings, such as effects of context window length or embedding vector size. They also compare embeddings produced using pre-training or by just training locally on the dataset of interest. Word embeddings are then evaluated by asking human subjects to rate the lists of the closest embedded words. However, no bootstrapping of the data is used, and stability metrics are only provided comparing different random initialisations of the neural network. As shown in Antoniak and Mimno (2018) and in our work, even small variations in the dataset could influence the stability of the embeddings and thus the associated closest word lists. Additionally, Rodriguez and Spirling (2022) evaluate their methods on data from the 102nd to 111th US Congresses, which contains 1.4M documents and 380M tokens, compared to our pre-training set of 26k documents and 160M tokens. This shows that their local training is comparable in the amounts of tokens we have in the pre-training data and greatly exceeds the amount of available data from the studied groups (motions from the Moderates and the Social Democrats). Therefore, using their methodology of directly training on all the available data was not feasible in our case.

Rodriguez, Spirling, and Stewart (2023) develop a context sensitive approach, studying how pre-trained embeddings in combination with regression techniques can be useful for investigating rare political terms that may change their meanings over context (compare Monroe, Colaresi, and Quinn (2008)). Their results indicate that if the purpose is to reveal rare terms, a specialized dataset for pre-training can be more useful than a more general one.

3.4. Implementation

Our setup (see Figure 2) is to start from one of the two pre-trained models and then fine-tune it on each of the different corpora we want to compare: motions from the Moderate party and the Social Democrats respectively.

For each party, and for each of the ten keywords (listed above), we extract the top 20 closest word embeddings, measured by cosine similarity. These lists can then be compared to highlight similarities and differences in word use between the parties. The overlaps (i.e. how many of the top 20 words are common between the two parties) and differences should reflect the nuances in the speakers' use of the term that each model manages to disambiguate.

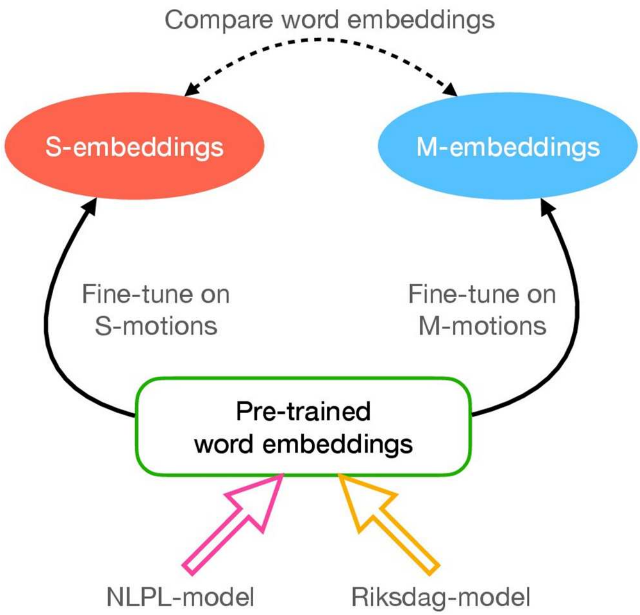

Figure 2. Fine-tuning word embeddings for Social Democrats and Moderates.

3.4.1. Training and fine-tuning word2vec

Our implementation of the word2vec model is based on the Gensim[4] library available in Python and the replication code can be found online[5] There are several hyperparameters that affect the model performance. For most we simply use the default settings (see replication code), for comparability the size of the embedding vectors we set it to 100, same as the pre-trained NLPL model.

3.4.1.1. Fine-tuning. Typically for word2vec models, when larger datasets are considered, only a few epochs are used, with the original paper (Mikolov et al. 2013) recommending 3, and the implemented library suggesting 5 as a default. Long fine-tuning of models is typically avoided as it could lead to *catastrophic forgetting* – a term used to describe the process of a neural network forgetting the previously learned patterns when trained on new data. This is however not an issue in our experiments where the aim is to amplify the patterns in the fine-tuning data for the Social Democrats and Moderate parties. As such, we were not worried about, nor did we see any signs of, loosing patterns from the pre-training data (e.g. terms not related to politics). Long fine-tuning was in fact beneficial and necessary: initial experiments with the default 5 training

[4]https://radimrehurek.com/gensim/
[5]https://github.com/dsaynova/word_embeddings_swedish_riksdag_motions

epochs showed large overlap in the top 20 word lists obtained for the parties. This indicates that the embeddings from the pre-training data which we started from had not changed much at all yet and still reflected the pre-training data. To remedy this, we needed to expose the model to much longer fine-tuning to amplify patterns from the respective parties' data.

To determine how well the model was adapted to our fine-tuning data, we measured the training loss on the word-from-context prediction training proxy task (as described in section 2.2). We choose the number of epochs with the elbow method – several epochs after the initial drop in training loss is observed. We found that 50 epochs allowed us to reach a plateau in the training loss for both pre-training approaches and both fine-tuning data sets – many more than the default values. We have opted for consistent number of epochs for both parties, so the number of epochs covers part of the plateau of the curve. This should be acceptable in our setting, as over-fitting to the motions data help amplify the signals for the patterns we want to discover.

3.4.1.2. Bootstrapping several runs. As there are random factors in neural network training (such as ordering of training data), results can be disproportionately influenced by a few samples (documents), especially when working with small to moderate sized datasets as we do (Antoniak and Mimno 2018). Therefore, we bootstrap our models by running them multiple times to aggregate results and get a measure of stability of the word embeddings and word closeness to keywords. This can be computationally expensive, why we run ten iterations of the following:

- Sample *n* motions with replacement, where *n* is the original size of the fine-tuning dataset.
- Fine-tune the pre-trained model on the extracted sample.
- Calculate cosine similarity for each of the keywords to all words in the vocabulary.

From all bootstrapped runs, we calculate the mean cosine similarity for each of the keywords to all words in the vocabulary and the standard deviation. Finally, we rank the vocabulary by mean cosine similarity and extract the top 20 words. We chose to focus on 20 words to have room for overlaps as well as distinctions between parties. The complete list of closest words can be found in the replication materials.

In the forthcoming analysis, we compare and discuss the word embedding's performance on different types of concepts: ideological words and issue words. While the number of studied terms is not sufficient to make statistical inferences of differences, a strength with our approach is that the terms are not randomly selected but based on voters' own descriptions of parties and previous related work. Thus, they represent a critical part of a party's image.

4. Results

For each of our ten keywords, we compare the 20 closest word embeddings for the respective parties on these dimensions measuring cosine similarity, its standard deviation and term frequency.

4.1. Common and different words between parties

In general, the parties have more similar vocabularies in some areas than in others (Table 3). More value laden terms like *solidarity* [solidaritet] have the least commonality between parties with both types of pre-training, while policy terms like *taxes* [skatt] and *crime* [brott] have the largest overlap. A tentative interpretation is that the more a concept orients towards an object such as for example taxes, the more the parties' vocabulary will converge, whereas the more the word is ideologically laden, the more differences between parties the word embedding approach will be able to deliver.

The amount of overlap is also related to how many synonyms and different forms a term has – as can be seen with the *crime* [brottslighet] example, where both other Swedish versions [brott, kriminalitet] appear in the lists along with the definite form of the word, e.g. [brottsligheten]. This will lead to some words naturally having bigger clusters of closely related terms to them and selecting the top 20 for all might not work equally well in detecting the full range of the nuances in usage.

The complete word lists of twenty-closest words for two of our keywords from the models are included in Appendix A: for the term *crime* in Table A1 for the NLPL model and in Table A3 for the Riksdag pre-trained model, and for the term *solidarity* in Table A2 for the NLPL model and Table A4 for the Riksdag model. The full resulting tables for all twenty-closest words for

Table 3. Number of overlapping (shared) words in the top 20 list between the two parties.

Term	Translation	NLPL	Riksdag
droger	drugs	9	7
skatt	taxes	13	12
säkerhet	state security	12	5
trygghet	social security	11	14
brott	crime	12	12
brottslighet	crime	6	7
kriminalitet	crime	8	7
jämlikhet	equality	8	11
solidaritet	solidarity	6	9
rättvisa	justice	10	9

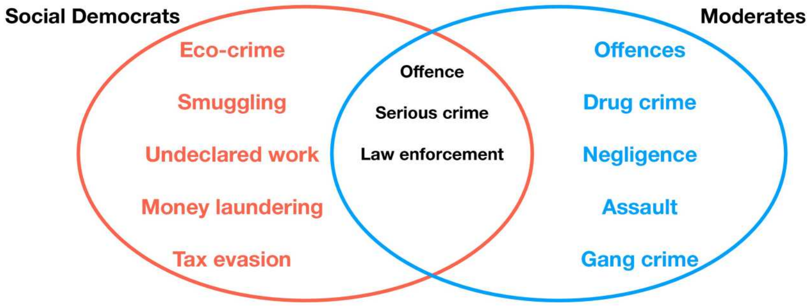

Figure 3. Word embeddings close to the term *crime* in the NLPL model trained on data from 1988 to 2020.

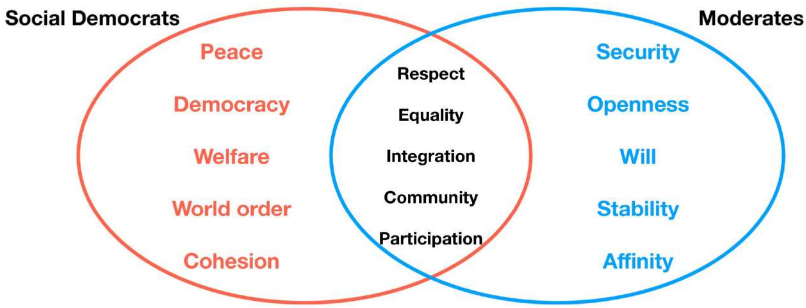

Figure 4. Word embeddings close to the term *solidarity* in the NLPL model trained on data from 1988 to 2020.

each term are available online, together with code for replicating the full results.[6] We observed that the NLPL model generally results in more words of political relevance, while the Riksdag model often tend to result in more words exemplifying the concept (e.g. lists of different types of crimes). To get an overview what kind of example words our models produce, a subset of politically meaningful words from the NLPL-wordlists are visualized in Figures 3 and 4, with focus on two different types of terms: one policy-oriented (*crime*) and one ideological (*solidarity*). We have selected the terms that illustrate politically relevant differences and similarities.

In the crime dimension, we find that Social Democrat parliamentarians' vocabulary focuses on crimes related to taxes, whereas the Moderates emphasize gang crime and assault during the studied period (see Figure 3). In the solidarity dimension, *peace* and *welfare* stand out as representative for Social Democrat vocabulary, whereas the Moderates associate the term with words such as *security* and *stability* (see Figure 4). These patterns

[6]https://github.com/dsaynova/word_embeddings_swedish_riksdag_motions

correspond well to the general tendencies of left-wing parties to focus on social and economic welfare, whereas the right-wing focuses more on security.

4.2. Reliability

Next, we turn to the question if we can trust these results as valid representations of the parties' vocabulary. We use a standard deviation measure: some terms have lower standard deviations in relation to the similarity score, meaning that their embeddings are more stable, and less affected by small changes to the data. As a guideline for measuring the reliability of the results, we can look at the number of terms that are more than one standard deviation above the similarity score of the twentieth word in the list (see Table 4). In Tables A1–A4 in the Appendix we report similarity scores along with the standard deviations. The example word should ideally have a score as close to 1 as possible, and low standard deviation. As such, the political scientist can be more confident in such results when drawing conclusions.

We see that a noticeable difference can be observed between terms, with *crime* [brott, brottslighet] and *taxes* [skatt] being the most stable for both parties. For Moderates the ideology related terms show the lowest reliability. Social Democrats also show instability in the ideology related terms, but additionally are less reliable for one of the *crime* terms [kriminalitet] and *drugs* [droger].

This is to an extent, but not linearly, aligned with the frequency of the studied keywords as can be seen in Figure 5. The squares represent term stability for the Social Democrats and the circles for the Moderates. The relationship is roughly logarithmic, but with a big variation. The terms *solidarity*, *equality* and *justice* are less stable than *drugs* and *crime* even though they have roughly the same frequency. Similarly, terms concerning security are less stable than those relating to crime. This is in line with similar work exploring

Table 4. Number of example words with a similarity score (from 0 to 1) more than one standard deviation higher than the score of the twentieth word in the list. A higher number means that the example words are closer to the term. Shown for both parties.

	Riksdag		NLPL	
	M	S	M	S
droger	13	6	13	9
skatt	10	12	14	11
säkerhet	9	7	11	12
trygghet	10	11	13	13
brott	10	14	12	12
brottslighet	10	10	8	9
kriminalitet	11	6	12	6
jämlikhet	4	14	4	10
solidaritet	11	9	8	8
rättvisa	8	12	7	9

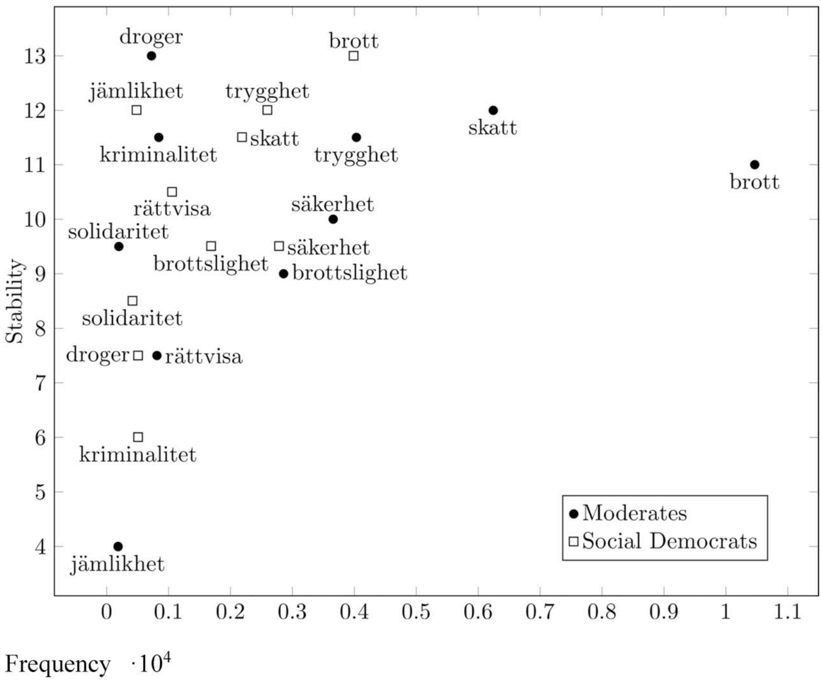

Figure 5. Stability comparison of terms. For how we define and calculate the stability refer to Section 4.2

embedding instability (Borah, Pratim Barman, and Awekar 2021; Wendlandt, Kummerfeld, and Mihalcea 2018), where a logarithmic relationship between frequency and stability is established and a large variability in the stability is also observed. Since frequency is not a perfect predictor for the difference in stability, especially in the low-frequency range, there could be other factors that contribute to this effect, like the type of words. One pattern we observe is that the ideological terms show less stability, which indicates that this method can have difficulties investigating discourse along more value laden dimensions. On the other hand, it is expected that policy issues that are referring to objects and facts show greater stability in this kind of data analysis. Word embeddings do not reveal political intentions in themselves, as these intentions can be latent and builds on a more complex combination of word expressions. Nevertheless, as we show, a word embeddings approach can still produce meaningful differences between parties.

Our findings also exemplify the importance of bootstrapping for estimating standard deviations of the results, especially when working with low signal and small to medium sized datasets. Previous works by Rodriguez and Spirling (2022) explore variability due to the stochastic nature of neural networks and Rodman (2020) considers variability due to re-sampling on a single term and its connection to five pre-selected topics. In our work, we identify that the reliability of the results can vary even within the same model setting depending on the terms that are explored. This variability can be influenced by term frequency and term type.

5. Discussion

From the word lists indicating the most stable and consistent word representations for the respective parties, we can observe some patterns that are more pronounced than others. Comparing the two different pre-training procedures, one main conclusion is that the external pre-training procedure produces words that indicate clearer ideological differences between the parties. We found that in the solidarity dimension, the Moderates emphasize security and stability, whereas the Social Democrats emphasize peace, democracy and welfare. This is a sign that representatives from the parties consistently have different associations vis-a-vis the same concept, depending on where they find themselves in the ideological space.

However, if the search is for political or technical jargon or trending words, the locally pre-trained embeddings can be valuable. In the solidarity dimension, for example, the term *polarization* comes up as an example word for both parties (see Table A4 in the Appendix). This suggests that the embeddings pre-trained on domain-specific data to some extents are better at picking up rare words. Still, in general, ideological markers do not appear as clearly in the domain-specific pre-trained embeddings as in the embeddings pre-trained on the larger general dataset.

Another important takeaway is that the word embeddings produced more stable results for the policy issues than for the ideological dimensions. One interpretation is that there are more variations in parliamentarians' way of discussing broader concepts and that the word embeddings computations then have a more difficult task to identify regularities. Still, the method can reveal useful and consistent patterns even for the more complex terms. Following the pre-training procedure including fine-tuning that we employ in the present study, party characteristic and word use can be revealed even from low signal political data.

6. Concluding remarks

This study elaborated and compared two ways of setting up a word embedding analysis for political science: one based on pre-training on a larger, external data source and the other on performing the pre-training on a smaller, domain-specific dataset sourced from the same domain as the study data: Swedish parliamentarians. We found that the pre-training procedure based on the larger, more general dataset, in combination with fine-tuning, produced the most reliable and useful results, distinguishing ideological differences between the parties. Thus, a specialized vocabulary could not fully compensate for dataset size, except in some cases or rare domain-specific terms, which is in line with previous results (Rodriguez, Spirling, and Stewart 2023). We also found some patterns that were independent from the pre-training procedure: starting from policy issue term such as *taxes* produced more reliable results but differentiated the parties less than ideological terms such as *equality* or *solidarity*. As a recommendation for practitioners, opting for off-the-shelf general pre-trained embeddings is a good starting point for most applications.

It is likely that this holds true for low signal data, which is the rule rather than the exception in proportional representation systems. In this study, we focused on the vocabularies of the two largest parties that compete for the Prime Minister position. This makes the present study case relatively similar to plurality system like the UK or the US, although the differences between blocs are less pronounced in our case. Parties that are smaller and less present in the debate could be even more difficult to evaluate using word embeddings, given that the data is smaller (Antoniak and Mimno 2018). However, as we have seen from our analysis, reliability of the results is also dependent on the frequency of the studied terms. Therefore, if the parties are issue-parties with a distinct profile, word embeddings analyses based on a small dataset could still tell something about their focus relative to other parties. In our study, the number of words in the dataset for the Social Democrats was smaller than for the Moderates, but Social Democrat emphasis on for example equality resulted in more stable results for them regarding that term.

Furthermore, we found that the differentiation and stability of the terms varied depending on frequencies in the data, and tentatively term type (issue or ideology). We anchored the choice of terms in voters' own descriptions of parties they could vote for and related the choice of concepts to other recent text-as-data studies (Rodman 2020; Rodriguez and Spirling 2022). Our strategy was to start from a number of central keywords and successively evaluate the 20 closest words that the word embedding iterations produced for both parties, combining stability metrics and knowledge about the political system. Forthcoming work should study more in-depth how well different word embedding methods perform studying different types of concepts. Our study indicates that studying terms that the parties interpret and use differently, such as *solidarity*, can require more iterations and expert evaluation than words that refer to issues such as *taxes*.

Our practices could also be further used for comparative political studies that aim to compare the differences (and similarities) in word use between competitor or coalition parties. One potential avenue for forthcoming research is to measure the distance in vocabulary between Prime Minister parties and a potential junior coalition party, to estimate its closeness to each of them. For the political scientist, it is still good news that an external pre-training set functions well, as it may be an extra effort to gather more local political data for training. Forthcoming analyses should also develop the relationship between unsupervised neural network techniques and supervised models, to obtain more knowledge about parties' vocabularies at a certain time.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program - Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation (2019.0209) and the Marcus and Amalia Wallenberg Foundation. We also want to thank Pasko Kisic Merino and discussants and participants at PolMeth Europe 2021 and APSA 2021. The computations and storage of data were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at Chalmers Centre for Computational Science and Engineering (C3SE), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

References

Antoniak, Maria, and David Mimno. 2018. "Evaluating the Stability of Embedding-Based Word Similarities." *Transactions of the Association for Computational Linguistics* 6:107–119. https://doi.org/10.1162/tacl_a_00008.

Bäck, Hanna, and Torbjörn Bergman. 2016. "The Parties in Government Formation." In *The Oxford Handbook of Swedish Politics*, edited by Jon Pierre, 206–226. Oxford: Oxford University Press.

Bäck, Hanna, Annika Fredén, and Emma A. Renström. 2022. "Legalize Cannabis? Effects of Party Cues on Attitudes to a Controversial Policy Proposal." *Journal of Elections, Public Opinion and Parties* 32 (2): 489–500. https://doi.org/10.1080/17457289.2021.1889570.

Bonafilia, Brian, Bastiaan Bruinsma, Denitsa Saynova, and Moa Johansson. 2023. "Sudden Semantic Shifts in Swedish NATO Discourse." In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, edited by Vishakh Padmakumar, Gisela Vallejo, and Yao Fu, 184–193. Toronto, Canada: Association for Computational Linguistics. https://aclanthology.org/2023.acl-srw.28.

Borah, Angana, Manash Pratim Barman, and Amit Awekar. 2021. "Are Word Embedding Methods Stable and Should We Care About It?" In *Proceedings of the 32nd ACM Conference on Hypertext and Social Media*, 45–55. New York, NY: Association for Computing Machinery.

Fernandes, Jorge M., Max Goplerud, and Miguel Won. 2019. "Legislative Bellwethers: The Role of Committee Membership in Parliamentary Debate." *Legislative Studies Quarterly* 44 (2): 307–343. https://doi.org/10.1111/lsq.12226.

Fredén, Annika, and Sverker Sikström. 2021. "Voters' Sympathies and Antipathies Studied by Quantitative Text Analysis: Evidence from a Two-Wave Panel Experiment in Sweden During Covid-19." In *Annual Midwest Political Science Association Conference*. Chicago: MPSA.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489–1501. Berlin, Germany: Association for Computational Linguistics.

Høyland, Bjorn, and Martin G. Søyland. 2019. "Electoral Reforms and Parliamentary Debates." *Legislative Studies Quarterly* 44 (4): 593–615. https://doi.org/10.1111/lsq.12237.

Mikolov, Tomás, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." In *1st International Conference on Learning Representations, ICLR 2013*.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16 (4): 372–403. https://doi.org/10.1093/pan/mpn018.

Osnabrügge, Moritz, Ash Elliot, and Massimo Morelli. 2021a. "Cross-Domain Topic Classification for Political Texts." *Political Analysis* 31 (1): 59–80. https://doi.org/10.1017/pan.2021.37.

Osnabrügge, Moritz, Sara B. Hobolt, and Toni Rodon. 2021b. "Playing to the Gallery: Emotive Rhetoric in Parliaments." *American Political Science Review* 115 (3): 885–899. https://doi.org/10.1017/S0003055421000356.

Proksch, Sven-Oliver, Will Lowe, Jens Wäckerle, and Stuart Soroka. 2019. "Multilingual Sentiment Analysis: A New Approach to Measuring Conflict in Legislative Speeches." *Legislative Studies Quarterly* 44 (1): 97–131. https://doi.org/10.1111/lsq.12218.

Rheault, Ludovic, Kaspar Beelen, Christopher Cochrane, and Graeme Hirst. 2016. "Measuring Emotion in Parliamentary Debates with Automated Textual Analysis." *PLoS ONE* 11 (12): e0168843.

Rheault, Ludovic, and Christopher Cochrane. 2020. "Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora." *Political Analysis* 28 (1): 112–133. https://doi.org/10.1017/pan.2019.26.

Rodman, Emma. 2020. "A Timely Intervention: Tracking the Changing Meanings of Political Concepts with Word Vectors." *Political Analysis* 28 (1): 87–111. https://doi.org/10.1017/pan.2019.23.

Rodriguez, Pedro, and Arthur Spirling. 2022. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *Journal of Politics* 84 (1): 101–115.

Rodriguez, Pedro L., Arthur Spirling, and Brandon M. Stewart. 2023. "Embedding Regression: Models for Context-Specific Description and Inference." *American Political Science Review* 117 (4): 1255–1274. https://doi.org/10.1017/S0003055422001228.

Tahmasebi, Nina. 2018. "A Study on Word2Vec on a Historical Swedish Newspaper Corpus." In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference, DHN 2018*, 25–37. Helsinki, Finland, March 7–9.

Wendlandt, Laura, Jonathan K. Kummerfeld, and Rada Mihalcea. 2018. "Factors Influencing the Surprising Instability of Word Embeddings." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2092–2102. New Orleans, LA: Association for Computational Linguistics.
Zeman, Daniel, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, et al. 2017. "CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies." In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, 1–19. Vancouver: Association for Computational Linguistics.

Appendix A. Top 20 closest words: Crime and Solidarity

The following tables contain the top 20 closest words for the keywords *crime* and *solidarity* for both Social Democrats and Moderates and for the models using the NLPL pre-trained embeddings and Riksdag pre-trained embeddings, respectively.

Table A1. Top 20 closest terms to brottslighet with NLPL model. Rows in bold indicate words that are more than one standard deviation away from the lowest score and are thus more reliable.

NLPL Moderates	Score(σ)	NLPL Social Democrats	Score(σ)
kriminalitet (crime)	**0.77(0.02)**	**brottsligheten (crime)**	**0.68(0.02)**
brottsligheten (crime)	**0.71(0.02)**	**kriminalitet (crime)**	**0.62(0.03)**
brott (crime)	**0.66(0.02)**	**skattefusk (tax fraud)**	**0.6(0.03)**
brotten (the crimes)	**0.63(0.02)**	**ekobrottslighet (eco – crime)**	**0.59(0.03)**
narkotikabrottslighet (drug crime)	**0.63(0.02)**	**smuggling (smuggling)**	**0.58(0.03)**
misskötsamhet (monitoring)	**0.62(0.03)**	svartjobb (black job)	0.57(0.06)
gängkriminalitet (gang crime)	0.6(0.04)	**ekobrott (economic crime)**	**0.57(0.05)**
våldsbrott (violent crime)	**0.6(0.02)**	**penningtvätt (money laundering)**	**0.57(0.04)**
narkotikasmuggling (drug smuggling)	0.59(0.03)	**brott (crime)**	**0.56(0.03)**
bedrägerier (fraud)	0.58(0.03)	ekobrottsligheten (eco – crime)	0.56(0.05)
vardagsbrottsligheten (everyday crime)	0.58(0.03)	**brottsbekämpning (law enforcement)**	**0.55(0.04)**
våldsbrottslighet (violent crime)	0.58(0.05)	kriminaliteten (crime)	0.55(0.05)
misshandel (abuse)	0.58(0.03)	påföljd (penalty)	0.53(0.03)
narkotikabrott (drug crime)	**0.57(0.01)**	verksamhet (operation)	0.53(0.02)
terrorism (terrorism)	0.57(0.04)	utsatthet (exposure)	0.53(0.02)
brottsbekämpning (law enforcement)	0.57(0.04)	grova (rough)	0.53(0.04)
människohandel (human trafficking)	0.57(0.02)	skattebrott (tax crime)	0.52(0.04)
grova (rough)	0.57(0.03)	korruption (corruption)	0.52(0.02)
brottet (the crime)	0.56(0.02)	våldsbrott (violent crime)	0.52(0.02)
sexualbrott (sexual crime)	0.56(0.02)	vinning (win)	0.52(0.02)

Table A2. Top 20 closest terms to solidaritet with NLPL model. Rows in bold indicate words that are more than one standard deviation away from the lowest score and are thus more reliable.

NLPL Moderates	Score(σ)	NLPL Social Democrats	Score(σ)
respekt (respect)	**0.56(0.03)**	**solidariteten (solidarity)**	**0.58(0.03)**
trygghet (security)	**0.54(0.02)**	**demokrati (democracy)**	**0.57(0.03)**
gemenskap (community)	**0.52(0.06)**	**fred (fred)**	**0.57(0.03)**
integration (integration)	**0.52(0.05)**	**integration (integration)**	**0.53(0.04)**
öppenhet (openness)	**0.5(0.02)**	**sammanhållning (cohesion)**	**0.53(0.03)**

(*Continued*)

Table A2. Continued.

NLPL Moderates	Score(σ)	NLPL Social Democrats	Score(σ)
vilja (want)	**0.5(0.03)**	**jämlikhet (equality)**	**0.51(0.03)**
sympati (sympathy)	**0.5(0.03)**	**respekt (respect)**	**0.51(0.02)**
försoning (reconciliation)	0.48(0.04)	gemenskap (community)	0.51(0.05)
empati (empathy)	0.48(0.03)	**välfärd (welfare)**	**0.5(0.03)**
säkerhet (security)	**0.48(0.02)**	samhällsmodell (social model)	0.5(0.04)
samhörighet (affinity)	0.48(0.04)	världsordning (world order)	0.49(0.06)
omtanke (thoughtfulness)	0.48(0.05)	humanism (humanism)	0.49(0.03)
sammanhållning (cohesion)	0.48(0.04)	medmänsklighet (humanity)	0.49(0.05)
medmänsklighet (humanity)	0.48(0.04)	samhällsutveckling (social development)	0.49(0.03)
delaktighet (participation)	0.47(0.03)	välfärdsmodell (welfare model)	0.48(0.03)
religionsfrihet (freedom of religion)	0.46(0.03)	människorna (the people)	0.48(0.03)
maktdelning (division of power)	0.46(0.05)	delaktighet (participation)	0.48(0.03)
jämställdhet (equality)	0.46(0.03)	människosyn (human view)	0.48(0.05)
stabilitet (stability)	0.46(0.03)	rättvisa (justice)	0.48(0.02)
tillit (considerate)	0.45(0.05)	folkhälsa (people's health)	0.47(0.04)

Table A3. Top 20 closest terms to brottslighet with Riksdag model. Rows in bold indicate words that are more than one standard deviation away from the lowest score and are thus more reliable.

Riksdag Moderates	Score(σ)	Riksdag Social Democrats	Score(σ)
kriminalitet (crime)	**0.74(0.01)**	**brottsligheten (crime)**	**0.63(0.03)**
narkotikabrottslighet (drug crime)	**0.68(0.04)**	**kriminalitet (crime)**	**0.61(0.04)**
gängkriminalitet (gang crime)	**0.67(0.03)**	**narkotikabrottslighet (drug crime)**	**0.61(0.07)**
brottsligheten (crime)	**0.67(0.02)**	**skattefusk (tax fraud)**	**0.6(0.03)**
våldsbrottslighet (violent crime)	**0.66(0.04)**	**miljöbrottslighet (environmental crime)**	**0.6(0.04)**
terroristbrottslighet (terrorist crime)	**0.64(0.02)**	**ekobrott (economic crime)**	**0.59(0.04)**
underrättelseverksamhet (intelligence activities)	**0.61(0.02)**	ekobrottslighet (eco − crime)	0.58(0.05)
brott (crime)	**0.6(0.01)**	**underrättelseverksamhet (intelligence activities)**	**0.58(0.02)**
narkotikasmuggling (drug smuggling)	**0.59(0.02)**	**narkotikabrottsligheten (drug crime)**	**0.57(0.03)**
misskötsamhet (monitoring)	0.59(0.03)	**terroristbrottslighet (terrorist crime)**	**0.57(0.03)**
grova (rough)	**0.58(0.01)**	smuggling (smuggling)	0.57(0.03)
kriminalitets (crime)	0.58(0.03)	kriminaliteten (crime)	0.57(0.04)
narkotikabrottsligheten (drug crime)	0.58(0.03)	**vinning (win)**	**0.56(0.02)**
våldsbrott (violent crime)	0.57(0.02)	narkotikasmuggling (drug smuggling)	0.56(0.03)
polisverksamhet (police operations)	0.57(0.03)	utsatthet (exposure)	0.56(0.04)
straffrihet (impunity)	0.57(0.05)	bedrägerier (fraud)	0.55(0.03)
narkotikahantering (drug)	0.56(0.03)	bedrägeri (fraud)	0.55(0.02)
brotts (crime)	0.56(0.04)	skattebrott (tax crime)	0.55(0.05)
terrorism (terrorism)	0.56(0.03)	ekobrottsligheten (eco − crime)	0.54(0.05)
människohandel (human trafficking)	0.56(0.03)	alkoholsmuggling (alcohol smuggling)	0.54(0.08)

Table A4. Top 20 closest terms to solidaritet with Riksdag model. Rows in bold indicate words that are more than one standard deviation away from the lowest score and are thus more reliable.

Riksdag Moderates	Score(σ)	Riksdag Social Democrats	Score(σ)
försoning (reconciliation)	**0.53(0.04)**	**fred (fred)**	**0.58(0.03)**
gemenskap (community)	**0.53(0.04)**	**solidariteten (solidarity)**	**0.55(0.04)**
respekt (respect)	**0.52(0.03)**	**demokrati (democracy)**	**0.55(0.02)**
integration (integration)	**0.5(0.04)**	**sammanhållning (cohesion)**	**0.55(0.03)**
demokrati (democracy)	**0.5(0.02)**	**humanism (humanism)**	**0.53(0.04)**
sympati (sympathy)	**0.5(0.04)**	**gemenskap (community)**	**0.53(0.02)**
religionsfrihet (freedom of religion)	**0.49(0.02)**	**integration (integration)**	**0.52(0.02)**
sammanhållning (cohesion)	**0.49(0.04)**	**världsordning (world order)**	**0.51(0.02)**
jämlikhet (equality)	**0.49(0.03)**	**fattigdom (poverty)**	**0.51(0.02)**
jämställdhet (equality)	**0.48(0.02)**	rättighetsperspektivet (the rights perspective)	0.5(0.02)
empati (empathy)	0.48(0.04)	människosyn (human view)	0.5(0.04)
öppenhet (openness)	**0.48(0.02)**	fattigdomsminskning (poverty reduction)	0.49(0.03)
samhörighet (affinity)	0.47(0.03)	värderingar (values)	0.49(0.03)
trygghet (security)	0.47(0.02)	demokratisering (democratization)	0.49(0.02)
medmänsklighet (humanity)	0.47(0.03)	samhällsutveckling (social development)	0.49(0.02)
värdighet (dignity)	0.46(0.04)	människovärdet (human dignity)	0.49(0.04)
tolerans (tolerance)	0.45(0.04)	försoning (reconciliation)	0.48(0.05)
demokratisering (democratization)	0.45(0.03)	tolerans (tolerance)	0.48(0.04)
polarisering (polarization)	0.45(0.04)	jämlikhet (equality)	0.48(0.02)
maktdelning (division of power)	0.45(0.04)	polarisering (polarization)	0.48(0.03)