Computational scaling of political positions from textual data using word embeddings

Patrick Schwabl[1]

Author Note

[1]Department of Media and Communication, LMU Munich

Correspondence concerning this article should be addressed to Patrick Schwabl, Department of Media and Communication, LMU Munich, Akademiestraße 7, 80799 Munich, Germany. Email: patrick.schwabl@ifkw.lmu.de

Introduction

Retrieving valid numerical estimates for positional stances toward politics has always been challenging in many disciplines. While social science has long used surveys and content analysis to this end, some methods try to scale positions from textual data automatedly. As one of these, the idea of representing words in a geometric space has been rediscovered (Mikolov et al., 2013; Osgood et al., 1957).

Generating valid estimates from textual data would save countless hours of coding. Connectedly, valid automation of such estimation would significantly increase the visible universe of analyzable textual data. It would also enable researchers to get fine-grained numerical values, perform algebraic calculations with them, and help standardize text as data usage across studies.

Three promising approaches

There are at least three proposals to these ends. The first one comes from research on finding bias in word embeddings. Bias in embeddings is present for concepts like gender (Bolukbasi et al., 2016) and ethnicity (Garg et al., 2018). These publications are often based on a measure that was coined, the word embedding association test (WEAT) by Caliskan et al. (2017).

For this test, a biased concept is defined as an n-dimensional subspace in a larger embedding space. Its layout is given by two groups of words that define the ends of the concept's subspace, for instance, [male, man, boy] and [female, woman, girl]. Calculating relative distances of this gender space with a word like nurse can validly estimate gender bias in particular terms. With `sweater`, there is also an R package to calculate WEAT scores (Chan, 2022). Although these ideas are promising for research in political communication and anything concerning text, they have not yet been used to estimate political positions.

Secondly, there is another approach called distributed dictionary representations (DDR). DDR "averages the representations of the words in a dictionary and uses that average to represent a given concept as a point in the semantic space." (Garten et al., 2018) Therefore, a researcher with domain knowledge identifies words presumed to define a construct. Thiele (2022) already implemented this idea in the R package `dictvectoR`.

The main difference to the WEAT-based idea is that DDR uses one dictionary instead of two and makes the transfer to a greater diversity of possible constructs.

Thirdly, there are contextualized word embeddings from transformer models, which are proven to have superior performance compared to static word embeddings on many tasks. (*GLUE Benchmark*, 2022; Vaswani et al., 2017) However, methods like DDR and WEAT are not easy to combine with contextualized embeddings. A recent study finds that "these methods fail to apply to contextualized embeddings due to their mutable nature." (Katsarou et al., 2022; see also Guo & Caliskan, 2021) Contextualized embeddings of single words change depending on the words they are surrounded by.

Fusing the three approaches

Katsarou et al. (2022) also hold a possible idea for bringing together the approaches. From a transformer model, they retrieve embeddings of the words "he" and "she" in 149 different sentences. Since these embeddings are contextualized, they change depending on the surrounding sentence.

One could transfer this to politics and any number of sentences containing the term "renewable energies" versus "fossil fuels." Combined with the idea of using a dictionary from DDR and association-based scores with two dictionaries like those given by the WEAT, one would get a method that can produce association scores of any embedded text with any dictionary in a non-static way. Doing so captures the different meanings that words describing the concept of interest have in different contexts. With this, one could measure the association of statically embedded speeches of a single politician with many hundred different meanings of "renewable energies" and "fossil fuels," gaining measurement validity compared to static word embeddings. This also transfers to broader political cleavages (Lipset & Rokkan, 1967), given proper dictionaries can be crafted.

That can also be approached from the other direction. This means that the axis of "renewable energy" – "fossil fuel" is to be held static. Then the contextually embedded texts of political actors can be associated with the static axis. "The insights from our exploration indicate that the use of a stable gender direction, even in a Transformer's mutable embedding space, can be a robust method to measure bias." (Katsarou et al., 2022) The axis is generated by aggregating different contextual word embeddings of different words in dictionaries.

So, either embedded political texts or embedded construct dictionaries can be mutable; the other should be static. Making both contextual would probably involve too many moving parts. However, to my knowledge, no study has tried this yet.

Setting up the first experiments

Data and dictionaries

I conducted computational experiments to show that the automatic scaling method proposed works in principle. Using speeches from the 19th German Bundestag (Richter et al., 2020), I devise four dictionaries to produce two political issue axes in a Word2Vec embedding space. Please note that the integration with transformers has not happened yet; for now, I used static word embeddings because they are faster to train but still sufficient to get a first impression of whether the method can work in principle.

I removed speeches that were shorter than 100 characters. I considered speeches from members of parliament, federal ministers, and the chancellor because these roles are not strictly separable. Lastly, I removed all punctuation and lowercased all speeches.

The space of a political dictionary serves as a touchstone against which Doc2Vec-embedded texts of politicians can be positioned (Figure 1). Note, that this is either a space, a plane, or an axis, depending on the number of dimensions involved. I will only refer to axes from here on.

I used the TensorFlow embedding projector (Smilkov et al., 2016) to dig into the embeddings (Figure 2). I crafted two custom axes in the projector using four dictionaries to represent the four ends of the axis. I identified the right words by reading some speeches and used nearest neighbors to snowball my way forward.

Crafting custom axes

The two axes I craft this way are a climate issue axis and a classical left-right axis. I use terms like "climate crisis" to define the end of the climate axis that indicates favoring stricter climate protection measures. Terms like "climate hysteria" determine the opposite end of this axis.

In spanning the left-right axis, I reason that left-leaning parties talk more about the extreme right with wording like "right-wing extremism" or "radical right." Therefore, these terms point to the end of the axis where the most left political position is. The other end of this axis is crafted by "left-wing ideological," "left-wing extremist," or similar terms (see Table 1 for the full German dictionaries).

I do not mean that only specific groups use specific words. That means if a right parliamentary group talks more about left extremists because it rails against the left, that is sufficient for discriminating the left from the right. It is about relative signal strengths, not about absolute ones.

I used principal component analysis on the Doc2Vec embedded politicians and their relative positions towards these two axes to make the 100-dimensional embeddings plottable in two-dimensional space.

Experimental results

What works so far – picking up strong signals

Some parties and politicians are more closely associated with the poles of issue spaces like renewable energy production and the left-right axis. Figure 3 plots both axes and the parliamentarians of the left-wing Greens and the right-wing AfD (Alternative for Germany). Interestingly, pronounced Green politicians like Anton Hofreiter, one of his party's very left-leaning figures, show up in the upper-right section. For the AfD, there is no such prominent politician in the lower left quadrant. However, more domain knowledge will help pin down the validity of such visualizations.

What also cannot be dismissed is that only a handful of politicians fall into the quadrants one would not expect them to be. The Greens should not appear in the lower left, and the AfD members should not appear in the upper-right quadrant. The visual separation is quite striking for a first experiment.

What does not work so far – picking up weak signals

Such striking visual separation is lacking when comparing the social democratic SPD with the Christian-conservative Union. However, it might not be the model missing signals, but instead, there being no or feeble signals, to begin with. SPD and Union have governed together since 2005, with one term of intermission from 2009 to 2013. They have been in governments together in many German states. The "social democratization" of the Union under Angela Merkel has been discussed in the German public and scientific literature. (Zolleis, 2015; Zolleis & Schmid, 2015) It might not be surprising that two parties making politics together for a long time are hard to distinguish in their wording for a model. Nonetheless, I am convinced that there is potential for word embeddings to also catch weak signals from alike parties such as SPD and Union.

What underscores this is the Pearson correlation of .50 ($p < .001$) of the left-right and climate scores for all members of parliament. Again, this coefficient is indicative that something is going on and that the method captures relevant signals.

Because of these promising findings, the next steps must be a theoretical, use-case-driven, methodological, and technical refinement.

Next steps

Apart from using contextualized word embeddings and bringing in relevant domain knowledge, as I already elaborated, I see three main next steps. Comments from the community on the next steps and anything concerning the paper are very welcome.

Standardizing dictionary development

First, I do not think how I crafted dictionaries here is the best way to do it. More theoretical work and qualitative evaluation are needed when applying the method to a new political conflict line. It was sufficient for a first impression, but there is room for improvement.

I want to get to a standardized way of crafting construct dictionaries. Combined with the method of deep reading and nearest neighbors, of help might also be a qualitative inductive categorization of how potential dictionary terms are used in the data. Using the nearest neighbor method, I was often stuck in a local cluster of useful terms. Qualitative methods can help identify terms that cannot be found by spatial proximity.

As a starter, I want to develop a handful of dictionaries defining the subspaces for single issues (Campbell et al., 1960) or societal cleavages and validate these and the resulting automation.

Get more and different political text data

Some 30,000 speeches from one electoral period in one country are insufficient to capture the universe of political positions. Therefore, I call for training models using more diverse data from the political domain, for instance, party manifestos, party websites, political news, social media data, laws and bills, texts from lobbying organizations or unions, Wikipedia articles, and international treaties.

I have already identified data sources, like the opted project (OPTED, 2022) or the comparative manifesto project (Volkens et al., 2021). It is not an easy dataset to craft, but I believe it will be some gigabytes in size and sufficient to train a transformer.

Assuring validity

To go beyond face validity for the scaled positions, I want to show ideas one could use to assess validity more thoroughly. A first step would be to look at whether the scaled positions are intrinsically coherent, whether they align with previous research, and whether they are complete measures in that respect (content validity).

Secondly, another step is to examine whether the results hold when comparing them against other measures (criterion validity). Comparing different automated scaling methods with each other, human coding, and expert surveys is a way to start. To date, the comparative manifesto project often serves as a gold standard since human evaluation of political positions from text is what research wants to capture.

Thirdly, construct validity should be tested by describing a research use case to which I will apply the method in future research.

Lastly, one could use a downstream task and take the scaled political positions to classify the party affiliation of politicians. If the scaled positions carry meaningful signals, they must be able to make such classifications correctly and, therefore, be a valuable representation of a construct.

References

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. https://arxiv.org/abs/1607.06520

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The american voter*. University of Chicago Press.

Chan, C. (2022). sweater: Speedy Word Embedding Association Test and Extras Using R. *Journal of Open Source Software*, *7*(72), 4036. https://doi.org/10.21105/joss.04036

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Garten, J., Hoover, J., Johnson, K. M., Boghrati, R., Iskiwitch, C., & Dehghani, M. (2018). Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis: Distributed dictionary representation. *Behavior Research Methods*, *50*(1), 344–361. https://doi.org/10.3758/s13428-017-0875-9

GLUE Benchmark. (2022). https://gluebenchmark.com/leaderboard

Guo, W., & Caliskan, A. (2021). Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 122–133. https://doi.org/10.1145/3461702.3462536

Katsarou, S., Rodríguez-Gálvez, B., & Shanahan, J. (2022). Measuring Gender Bias in Contextualized Embeddings. *AAAI Workshop on Artificial Intelligence with Biased or Scarce Data (AIBSD)*, 3. https://doi.org/10.3390/cmsf2022003003

Lipset, S. M., & Rokkan, S. (Hrsg.). (1967). *Cleavage structures, party systems, and voter alignments. An introduction*. Free Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ArXiv:1301.3781 [Cs]*. http://arxiv.org/abs/1301.3781

OPTED. (2022). *OPTED – Observatory for Political Texts in European Democracies: A European research infrastructure*. https://www.opted.eu/

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* (S. 342). Univer. Illinois Press.

Richter, F., Koch, P., Franke, O., Kraus, J., Kuruc, F., Thiem, A., Högerl, J., Heine, S., & Schöps, K. (2020). *Open Discourse* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/FIKIBO

Smilkov, D., Thorat, N., Nicholson, C., Reif, E., Viégas, F. B., & Wattenberg, M. (2016). *Embedding Projector: Interactive Visualization and Interpretation of Embeddings* (arXiv:1611.05469). arXiv. http://arxiv.org/abs/1611.05469

Thiele, D. (2022). *DictvectoR: Word vectors for dictionaries*. https://thieled.github.io/dictvectoR/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *ArXiv:1706.03762 [Cs]*. http://arxiv.org/abs/1706.03762

Volkens, A., Burst, T., Krause, W., Lehmann, P., Matthieß, T., Regel, S., Weßels, B., Zehnter, L., & Wissenschaftszentrum Berlin Für Sozialforschung (WZB). (2021).

Manifesto Project Dataset [Data set]. Manifesto Project. https://doi.org/10.25522/MANIFESTO.MPDS.2021A

Zolleis, U. (2015). Auf die Kanzlerin kommt es an. In R. Zohlnhöfer & T. Saalfeld (Hrsg.), *Politik im Schatten der Krise* (S. 73–91). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-05213-3_4

Zolleis, U., & Schmid, J. (2015). Die CDU unter Angela Merkel – der neue Kanzlerwahlverein? In O. Niedermayer (Hrsg.), *Die Parteien nach der Bundestagswahl 2013* (S. 25–48). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-02852-7_2

Tables & Figures

Figure 1: Schematic scratch of embedding associations between groups of seed dictionaries (defining a migration axis) and speeches of two Bundestags factions.

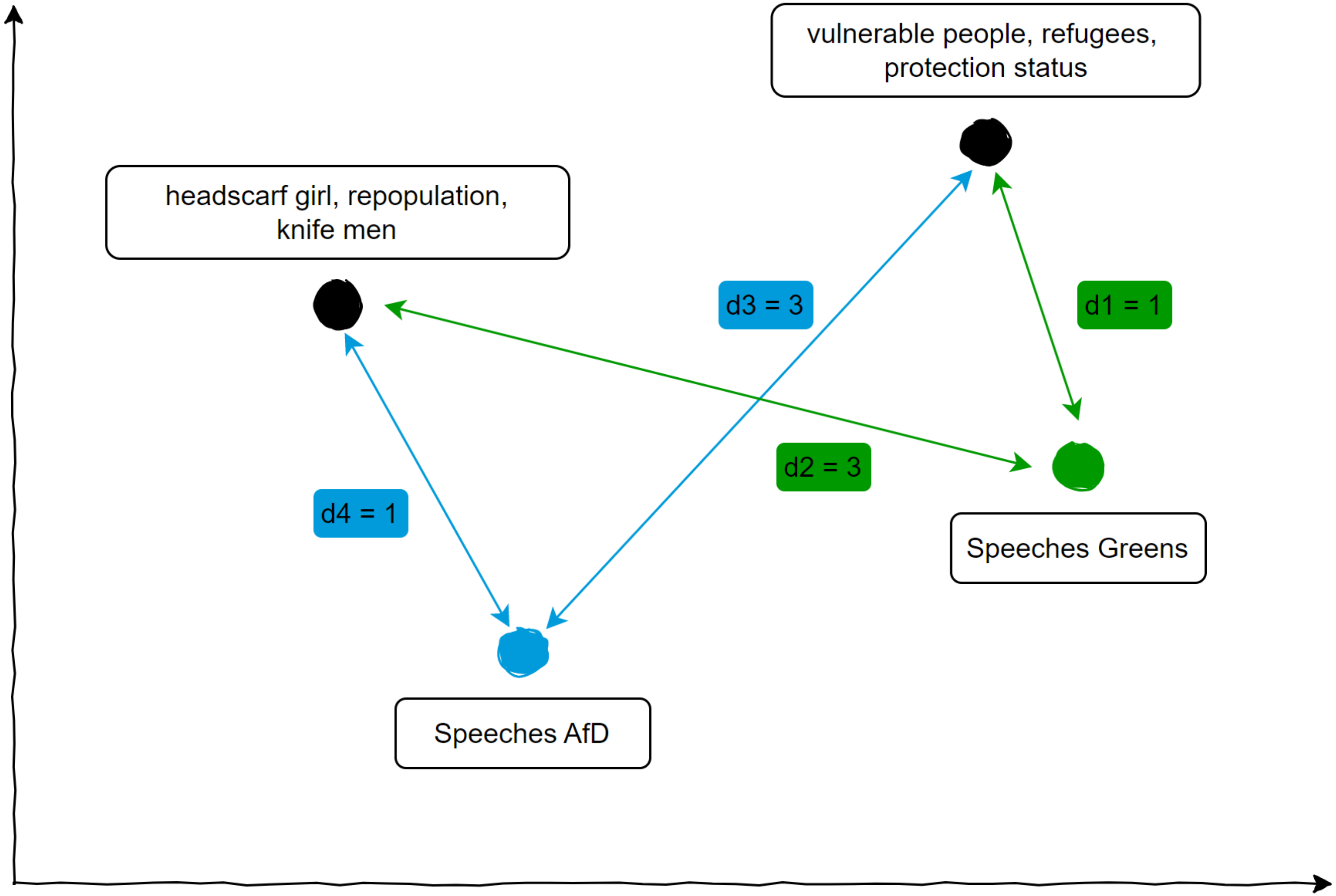

Figure 2: Tensorflow embedding projector with politician embeddings.

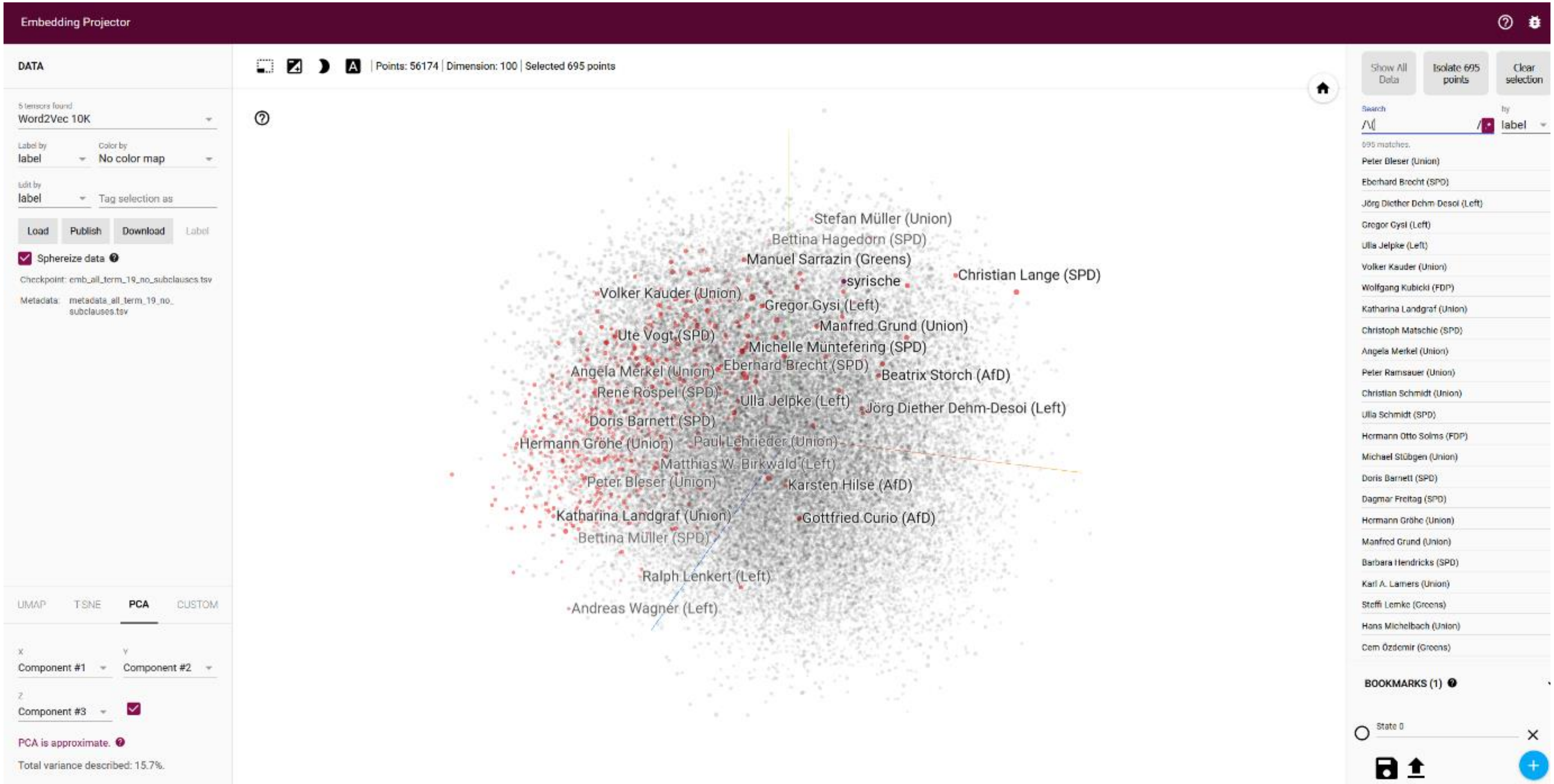

Figure 3: Left-wing Green and right-wing AfD politicians in 2d space (left-right axis vs. climate axis).

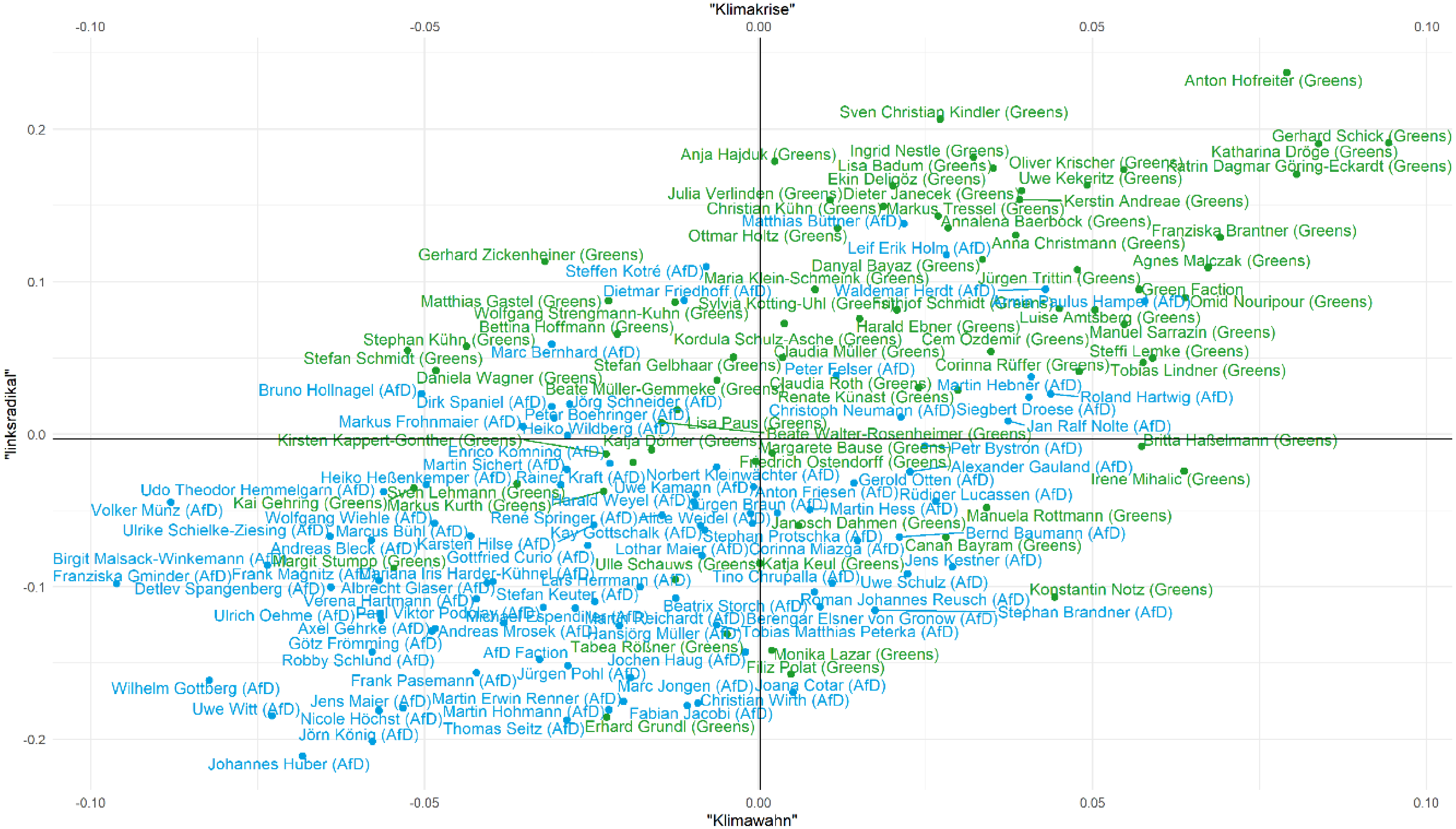

Figure 4: Conservative Union and social democratic politicians in 2d space (left-right axis vs. climate axis).

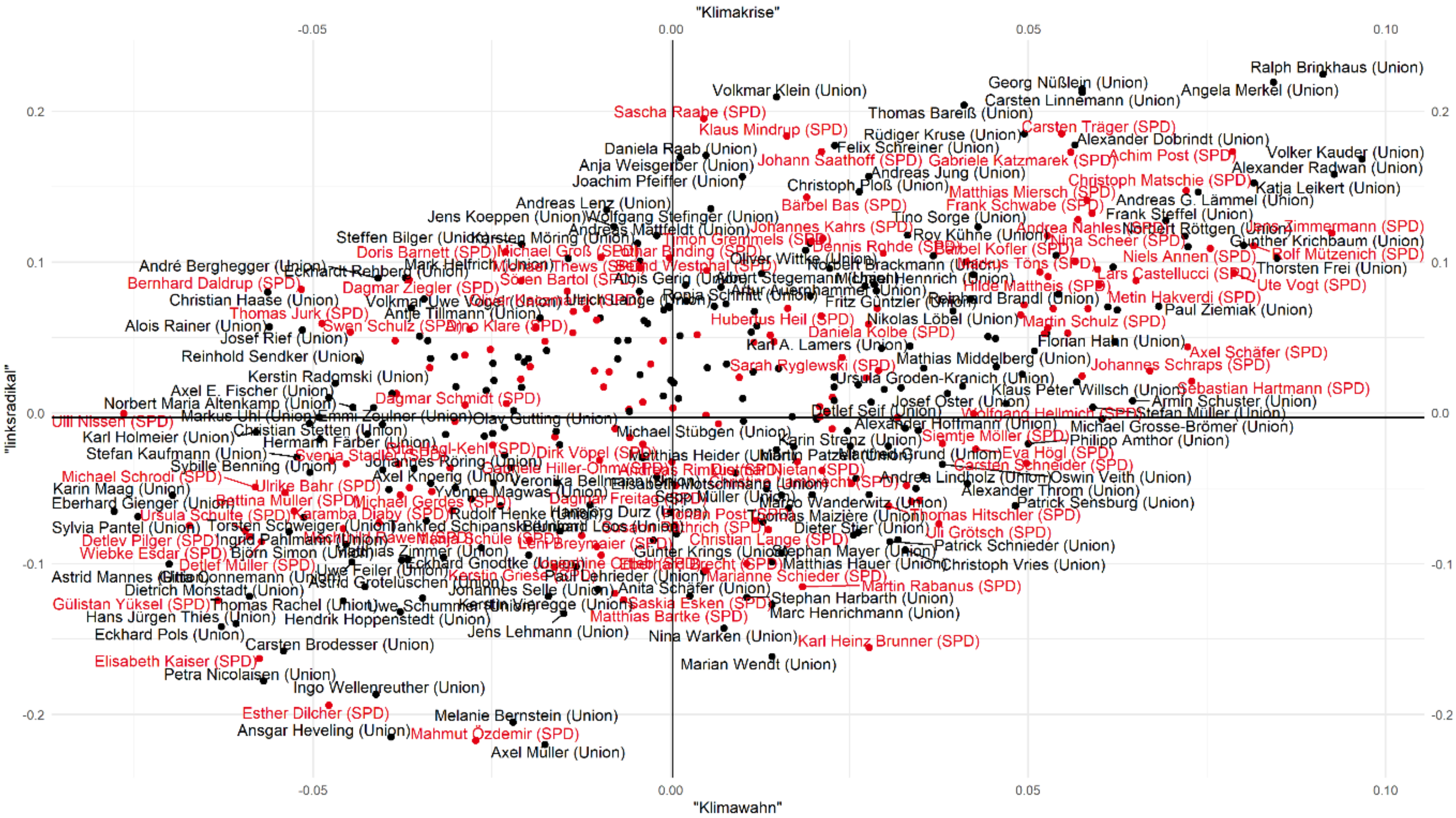

Table 1: German seed dictionaries.

Pro stricter climate measures	Contra stricter climate measures	Left-wing	Right-wing
klimakrise	klimahysterie	rechtsextreme	linksideologischen
klimakatastrophe	klimahysteriker	rechtsextremistischen	linksgrünen
klimagerechtigkeit	klimaideologie	rechtsextrem	linksextremistische
klimarettung	ökosozialismus	rechtsextremismus	linksextremistischen
klimaleugner	klimawahn	rechtsextremen	linksextremismus
klimanotstand	klimareligion	rechtsextremer	linksideologische
		rechtsextremistischer	linksextremistischer
		rechtsextremistische	linksextremen
		rechtsextremist	linksextremisten
		rechtsextremistisch	linksradikalen
		rechtsextremisten	linksextreme
		rechtsextremem	linksradikale
		rechtsextremistisches	linksextremistisch
		rechtspopulistischen	linksextrem
		rechtspopulismus	linksterroristen
		rechtspopulisten	linksgrüne
		rechtspopulistische	linksterrorismus
		rechtsradikalismus	linkspopulismus
		rechtsradikalen	linksgrüner
		rechtsradikal	linksradikal
		rechtsradikaler	antifa
		rechtsradikale	
		faschisten	
		neonazis	
		neonazistische	
		faschisten	
		flügel	

Note: The Flügel is an organization within the Alternative für Deutschland and counts as a far-right organization within the right-wing AfD. The left-right axis generally has more words because it relies on verbs, for which I included the occurring inflected versions.