The ties that bind.
Text similarities and the conditional diffusion of party policies.

Supporting information

June 2020

Table of Contents

Supporting information

The supplemental information contains six parts: In the first part, we give background information for the machine translation and our validation approach. In the second part, we report descriptive statistics for the (in-)dependent variable(s) used in the regression analysis. The third part contains the regression tables of the models corresponding to Figure 2 in the main text followed by alternative model specifications. We examine the relationship of text similarity and left-right party positions in the fifth part. Finally, in the sixth section we discuss Jaccard similarity as an alternative dependent variable. More comprehensive information is available as R Markdown/HTML files at doi:10.7910/DVN/ZJGHKK.

Appendix A: Making use of machine translation

Although desirable, having professional human translations of each party manifesto is virtually impossible given the amount of text and the costs it would require. While machine translation is far from being perfect it is much cheaper and faster. Furthermore, it has matured to a degree that it becomes a feasible option for answering substantial research questions involving cross-country comparisons and multilingual text corpora (Lucas et al. 2015; Reber 2018; Vries et al. 2018). What is more, these studies suggest that working with translated document-feature matrices (DFM) instead of having full-text translations is sufficient for analytical purposes in most cases. As this dramatically reduces the amount of text being translated by up to 95 percent, it becomes all the more interesting for small-scale projects.

But even then, machine translation is a matter of available resources as service providers charge a fee based on the number of characters that are translated. Full-text translations of the entire Manifesto Corpus, for example, would easily exceed $10,000. This would expand the space for more fine-grained analyses and hopefully they become available in the near future. Given limited resources though, decisions have to be made which ultimately reduce the amount of text that needs translating – yet leaving as many options for later analyses as possible.

Limiting the amount of text

The Manifesto Corpus (Version 2018-2) currently does not match with the main dataset's scope. This sets limits for our analyses. For most Eastern European countries, manifestos are available since the early 2000s, with a few exemptions starting in the mid-1990s. Excluding the Eastern and non-European countries allows for a much longer time perspective, though. The analysis of party policy diffusion only recently gained momentum with a focus on established Western European democracies (Böhmelt et al. 2016; Böhmelt et al. 2017). Thus, we limit our analysis to developed democracies and highly industrialized countries that (a) have a long history of competitive and fair elections, and (b) relatively stable party systems. We exclude Japan because it "really is a one-off case […], since it is the only instance of a country of non-European antecedents to become an advanced capitalist democracy" (Castles 1998, 9). For most Western countries, manifestos are available since the early 1960s – the Netherlands, France and Germany being the sole exceptions. In addition, Iceland, Luxembourg and Greece are ill-covered starting as late as the 2000s and mid-1990s, respectively. Focusing on the textual level, we are therefore currently bound by data availability. Besides, solely looking at Western countries efficaciously reduces the number of language pairs that need to be considered. Given that the more data, the better the performance, machine translation algorithms are less trained for rather uncommon languages. Because machine translation per se is not perfect, this would introduce noise to our data. Thus said, we limit the translation to 19 Western (European) countries. This reduces the number of language pairs that need to be translated to English to twelve. For one, all manifestos from the English-speaking family can be taken "as is". Second (and luckily in terms of resources), in some countries, manifestos are issued in common languages (e.g. Swiss manifestos are either published in French, German or Italian). This leaves us with the following languages that need to be considered for translation: Catalan, Danish, Dutch, Finnish, French, Galician, German, Italian, Norwegian, Portuguese, Spanish, and Swedish.

English will serve as the *tertium comparationis* because it is virtually the *lingua franca* of computing. For this reason, machine translation algorithms are mostly and best trained for translating to and from English, especially when large parallel corpora exist (Vries et al. 2018, 418–19). Rumor has it that English is also used as a "bridge" language when there is a lack of such corpora (Lucas et al. 2015, 269). We are using one of the most popular machine translation providers – Google Translate – and their API to do the translation. Google Translate offers all required languages and all are supported by their Neural Machine Translation Model for more accurate results. We thereby circumvent potential bias arising from using different providers that apply different models. Comparing Google Translate and DeepL, Reber (2018, 12–13) indeed concludes: "[I]t is safe to say that the choice of translation service plays a minor role. More important is the choice of the translation method […]".

Doing a DFM translation instead of a full-text translation reduces the amount of text by up to 95 percent. As Google and other service providers charge a fee based on the number of characters, this can make a large difference especially for smaller projects with limited resources. Typical steps in quantitative text analysis affect what will be eventually translated. This encompasses first and foremost tokenization ("bag-of-words"), and then lowercasing, removal of stopwords, numbers or characters, removal of very sparse or very frequent terms, lemmatization, and stemming (Grimmer and Stewart 2013; Welbers et al. 2017). A notable reduction in the number of features is achieved by stemming words. Stemming converts inflected forms of words into their base forms on a rule-based algorithm (Welbers et al. 2017, 251). Going one step further, lemmatization would replace the word with its morphological root. It certainly makes sense to stem or lemmatize for a given research question. For the translation, however, even stemming would entail too much loss of information. At the same time, stemming leads to negligible decreases in the number of characters which would not outweigh the loss of information. Thus said, we only tokenize, remove punctuation, numbers and symbols, and lowercase all words. Similar to Reber (2018, 5), we thereby retain as much information as possible for the translation.

Validating the translation

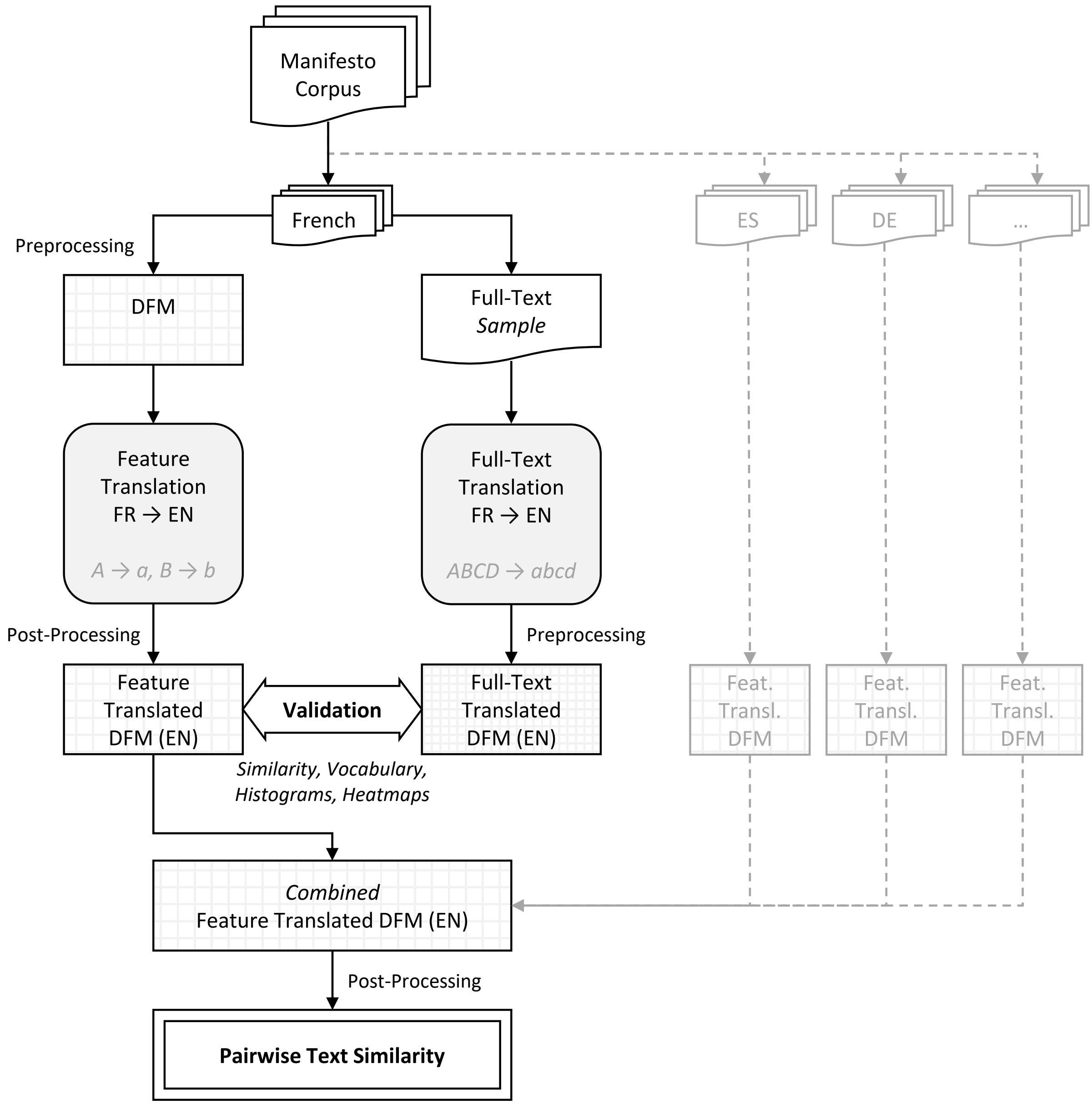
According to Grimmer and Stewart (2013, 271), Principle Four of automated text analysis is to "validate, validate, validate". Likewise, Vries et al. (2018, 419) argue for evaluating "the implications of machine translation for bag-of-words methods". Our dependent variable in the causal analysis – pairwise text similarity of party manifestos – rests on DFMs. Yet, when constructing a DFM we disregard the order in which words appear thereby losing context ("bag-of-words approach").

Neural networks are trained to take context into account – one reason why machine translation matured at all. However, when translating only the features, we run the risk of mistranslations as in the early days of phrase-based models. For this reason, we compare and validate the translation of each DFM with a DFM derived from a sample of full-text translations. Note that we are not validating machine translation per se as previous research has shown that full-text machine translation has caught up to human-translated texts (Le and Schuster 2016; Lotz and van Rensburg 2014; Lucas et al. 2015; Vries et al. 2018). If human translations were the "gold standard", we instead compare the "silver standard" of full-text machine translation to the "bronze standard" of DFM-translations. To our knowledge, only Reber (2018) provides first evidence of the suitability of such an approach.

Approach

Our approach is visualized in the following flow chart. This way, we can assess whether there are any (systematic) deviances and whether too much information is lost in the DFM translation that would thwart cross-lingual comparisons. If both DFMs turn out to be quite similar, we can be sure that our dependent variable rests on solid ground – an essential cornerstone for comparing party manifestos across countries and time.

Figure 1: Flowchart of translation approach



We start with the Manifesto Corpus and separate the documents according to the language they are written in. For each subcorpus, i.e. all documents of one language, we take two "paths":

- **Path One – Feature translated DFM**: In the first path, we preprocess the subcorpus, i.e. we tokenize, remove punctuation, numbers and symbols, and lowercase. We then extract the features, translate them and replace them in their DFM. In the following, we denote the DFM that is constructed in the original language prior to the translation and post-processed afterwards "feature translated DFM". Post-processing is necessary for several reasons. For one, some languages are well known for compound words (Lucas et al. 2015, 258). Often, their translation results in n-grams (e.g. the German "Sozialpolitik" is translated as "social policy"). For this reason, we post-process the DFM following Reber's (2018, 5) advice to split them into unigrams and remove duplicate terms. This is justified because "[i]n practice, for common tasks […] n-grams do little to enhance performance" (Grimmer and Stewart 2013, 272). Secondly, at this stage we remove English stopwords. By definition, stopwords do not contain topical content (they rather serve grammatical functions), but they affect the word distribution and subsequent measures (Vries et al. 2018, 421). In our case, for example, we would overestimate the similarity of a feature-translated document with its full-text companion. We disregard other typical steps like stemming or removing sparse or frequent terms. They make sense, for example, when fitting topic models as one is interested in carving out the content of topics and their word associations. At this stage, however, we are "simply" interested in the equality of both translation methods.
- **Path Two – Full-text translated DFM**: The second path involves full-text translations of entire documents. Given limited resources though, we draw a random sample from each subcorpus proportional to a language's share. In sum, we translate 260 manifestos (≈20 percent) as full-text. In the French example, we translate 30 party manifestos. We then construct the DFM applying the same steps as for the feature translated DFM, i.e. we tokenize, remove punctuation, numbers, symbols, English stopwords, and lowercase. Splitting compound words is not necessary in this case as their English equivalent is already separated when tokenizing. Likewise, we neither stem nor remove uncommon or frequent terms to retain as much information as possible about the translation when comparing both DFMs. In the following, we denote the DFM that is constructed and preprocessed after the translation "full-text translated DFM".

Having both the feature translated DFM and their full-text translated counterpart at hand, we can compare and validate both translation methods. We start by summarizing the comparison of the very same documents from both DFMs to see how the feature translation fares compared to a full-text translation. Afterwards, we assess whether the feature translated DFMs are able to detect the same patterns found in the full-text sample. If so, we can be sure that we are able to uncover patterns of text similarity across parties and countries as well.

Comparing both translation methods

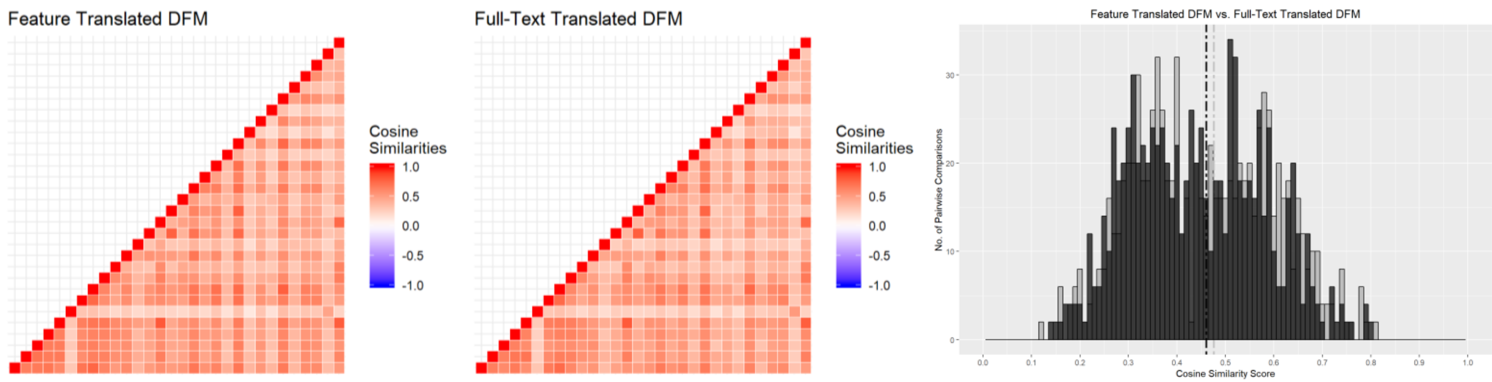
To evaluate the equivalence, we rely on two similarity measures – cosine similarity and the Jaccard coefficient. For one, cosine similarity can be regarded as a baseline model (Bär et al. 2015, 16). It is often used, not least because the resulting metric is on a familiar scale from 0 to 1. In addition, it has the advantage of being independent from document length (Huang 2008, 51). Applying the Jaccard coefficient is less common, but well-suited for our validation. In short, it quantifies the overlap of two vectors (here simply the features) and ranges from 0 to 1 as well (Huang 2008, 52). The drawback is that it does not take into account the frequency of terms, plus it is sensitive to document length. Thus, the Jaccard coefficient usually gives lower values than the cosine similarity. Used in conjunction, however, both are appropriate to assess the equivalence of the two translation methods. Cosine similarity and the Jaccard coefficient are implemented in common text analysis tools such as *quanteda* (Benoit et al. 2018), which we use for our analyses. Apart from document similarity, we also look at the vocabulary of both DFMs to check their overlap, and consider the degree of non-translation (Lotz and van Rensburg 2014, 250) in the feature translated DFM deriving from words that Google Translate could not match or other errors. This way, we get a clear picture of the accuracy and the shortcomings of Path One. Furthermore, we estimate all pairwise similarities among the sampled full-texts and compare the "correlation" matrix to the same matrix derived from the feature translated DFM by estimating the correlation between both matrices and inspecting heatmaps and histograms. Figure 2 gives an example for the French subcorpus. Similar graphs and a comparison of the Top 50 features of each language are available as an R Markdown/HTML file in the Dataverse.

Table 1 displays summary statistics for the comparison of the same documents – once from the feature translated DFM and once from the full-text translated DFM. The second and third column contains the number of manifestos for each subcorpus and how many documents were randomly sampled. The fourth and fifth column show the number of unique types in both DFMs. The sixth to the ninth column show the mean and standard deviation of the pairwise similarity. The tenth column shows the number of non-translated features per language. Finally, the last column shows Pearson's correlation coefficient for both matrices of pairwise similarities.

Table 1: Descriptive statistics for comparing feature vs. full-text translations

Language	No. of manifestos	Sampled manifestos	Types full-text translated DFM	Types feature translated DFM	Cosine Mean	Cosine SD	Jaccard Mean	Jaccard SD	Non-translated Features	Matrix correlation
Catalan	12	2	10,697	9,617	0.72	0.12	0.57	0.00	2,200	1.00
Danish	175	35	21,932	21,154	0.76	0.07	0.45	0.04	2,439	0.96
Dutch	199	39	90,581	84,305	0.78	0.10	0.55	0.03	6,765	0.93
Finnish	81	16	12,586	13,102	0.65	0.10	0.37	0.07	1,369	0.96
French	152	30	55,633	50,935	0.79	0.04	0.51	0.03	9,203	0.98
Galician	5	2	4,829	4,713	0.62	0.19	0.55	0.02	692	1.00
German	237	44	88,508	85,200	0.80	0.07	0.45	0.04	12,039	0.95
Italian	100	20	35,675	34,098	0.75	0.11	0.53	0.02	3,672	0.93
Norwegian	95	19	53,926	52,349	0.71	0.13	0.48	0.08	7,217	0.77
Portuguese	66	13	29,537	27,880	0.80	0.04	0.54	0.03	2,825	0.99
Spanish	81	16	58,338	54,704	0.78	0.05	0.56	0.04	11,374	0.93
Swedish	119	24	13,116	12,238	0.78	0.05	0.45	0.03	1,082	0.96
TOTAL	1,322	260	475,358	450,295	0.74	0.09	0.50	0.04	60,877	0.95

Figure 2: Example of heatmap and histogram for pattern of pairwise similarities in French subcorpus



In general, we find a large overlap although the vocabulary is slightly reduced compared to the full-texts. The mean cosine similarity is close to or even above the values that Reber (2018, 10) reports. This leaves us in good company. Apparently, information is lost when translating only the features and Google surprisingly has trouble translating some terms at all. Furthermore, it becomes obvious that the neural network is better trained on “larger” languages. Yet, the overall picture shows that the feature translated DFMs are sufficiently accurate and equivalent.

Conclusion

Comparing party manifestos across time and space requires a translation from the original language into a common “analysis” language. We chose English as the target language as it is the *lingua franca* of computing and Google’s neural network is best trained for translating to and from English. Having a vast amount of text, applying machine translation is inevitable given the costs for professional human translation. But even then, machine translation requires resources. In order to reduce the amount of text, translating only the features of a DFM seems to be a cost-effective solution. Yet, when translating only single terms one runs the risk of erroneous translations as in the early days of phrase-based models. Although Lucas et al. (2015) and Reber (2018) showed that – in principle – a feature translation is sufficient, Grimmer and Stewart (2013) remind us to validate automated text methods. For this reason, we proposed an intermediary approach by comparing the feature translated DFM to a DFM based on a random sample of full-text translations. Comparing both translations shows that the accuracy varies between languages, though in general we find a large overlap in terms of features. Interestingly, languages with a small number of randomly sampled full-texts have lower mean cosine similarities. This leads us to suspect that we (slightly) underestimate the equivalency of both translation methods in these cases. We attribute the observed differences between “small” languages such as Finnish and world languages such as Spanish or French to the peculiarities of the respective languages and the quality of Google Translate’s neural network. Furthermore, we do not find any systematic difference for one language. Thus, it is fair to assume that machine translation is subject to a small amount of – randomly distributed – errors. The most important fact, though, is that the feature translated DFM are able to detect the same pattern of pairwise similarities observable in the full-text translated DFM. This puts confidence in the feature translated DFMs being sufficiently accurate for estimating the pairwise similarity of party manifestos across time and space. We agree with Reber (2018, 13) and Lucas et al. (2015, 270) that a full-text translation is always more desirable. Especially for pilot studies and smaller projects though, even the “bronze standard” of machine translated features provides a very cost-effective approach for answering substantial research question involving multilingual corpora.

Appendix B: Descriptive statistics

Table 2: Descriptive statistics for numeric variables

Variable	Obs.	Mean	SD	Median	Min.	Max.
Cosine similarity	105,575	0.27	0.13	0.25	0.00	1.00
Past vote gain/loss$_{sender}$	105,575	-0.27	4.56	-0.20	-28.00	20.60
Previous vote gain/loss$_{sender}$	105,575	-0.14	4.31	-0.10	-27.00	20.30
Previous vote gain/loss$_{receiver}$	105,575	-0.10	4.40	-0.10	-27.00	20.30

Note: For cross-border diffusion the **past** vote gain/loss of the sender is relevant (i.e. t_0 to t_{-1}), for the domestic context the **previous** gain/loss applies (i.e. t_{-1} to t_{-2}).

Table 3: Frequencies of factor variables

Variable	Frequency	% valid	% valid cumul.
"Recycling"			
0	104,426	98.91	98.91
1	1,149	1.09	100.00
Competitors			
0	100,209	94.92	94.92
1	5,366	5.08	100.00
From Governments			
0	78,961	74.79	74.79
1	26,614	25.21	100.00
Among Governments			
0	83,008	78.62	78.62
1	22,567	21.38	100.00
EP factions			
0	99,976	94.70	94.70
1	5,599	5.30	100.00
Transnational party organizations			
0	97,401	92.26	92.26
1	8,174	7.74	100.00
Family of nations			
0	73,791	69.89	69.89
1	31,784	30.11	100.00

Appendix C: Regression models for Figure 2

The following tables contain the corresponding regression models on which Figure 2 is based.

Table 4: Interaction models sender attributes

	"Recycling"	Competitors	From gov.	Among gov.	EP factions	TPOs	Fam. of nations
"Recycling"	0.406***	0.406***	0.406***	0.405***	0.400***	0.399***	0.359***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Competitors	0.228***	0.228***	0.228***	0.227***	0.228***	0.228***	0.180***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments			-0.002*				
			(0.001)				
Among governments				0.015***			
				(0.001)			
EP factions					0.017***		
					(0.001)		
Transnational party organizations						0.019***	
						(0.001)	
Family of nations							0.064***
							(0.001)
Previous gain/loss$_{sender}$	0.000*	0.000					
	(0.000)	(0.000)					
Past vote gain/loss$_{semder}$			-0.000***	0.000*	-0.000***	-0.000***	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
*Previous gain/loss$_{sender}$ * Tie*	0.001	0.001***					
	(0.001)	(0.000)					
*Past vote gain/loss$_{semder}$ * Tie*			0.001***	-0.001***	0.000	0.000*	-0.000*
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	0.270***	0.270***	0.270***	0.267***	0.269***	0.268***	0.256***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: elecid.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: elecid.j (Intercept)	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Var: Residual	0.007	0.007	0.007	0.007	0.007	0.007	0.006
Num. groups elecid.i	290	290	290	290	290	290	290
Num. groups elecid.j	290	290	290	290	290	290	290
AIC	-218,966	-218,975	-219,012	-219,477	-219,136	-219,317	-229,160
BIC	-218,842	-218,851	-218,878	-219,344	-219,002	-219,183	-229,026
LL	109,496	109,501	109,520	109,753	109,582	109,673	114,594
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel models with non-hierarchical random intercepts for elections; decade FEs included but not shown; levels of significance: *** p < 0.001, ** p < 0.01, * p < 0.05.

Table 5: Interaction models receiver attributes

	"Recycling"	Competitors	From gov.	Among gov.	EP factions	TPOs	Fam. of nations
"Recycling"	0.407***	0.406***	0.406***	0.405***	0.400***	0.399***	0.359***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Competitors	0.228***	0.228***	0.228***	0.227***	0.228***	0.228***	0.180***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments			-0.001*				
			(0.001)				
Among governments				0.016***			
				(0.001)			
EP factions					0.017***		
					(0.001)		
Transnational party organizations						0.019***	
						(0.001)	
Family of nations							0.064***
							(0.001)
Previous gain/loss$_{receiver}$	0.000	0.000	0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
*Previous gain/loss$_{receiver}$ * Tie*	0.001	0.000	0.000	0.000*	-0.000	-0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept	0.270***	0.270***	0.271***	0.267***	0.269***	0.269***	0.256***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: elecid.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: elecid.j (Intercept)	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Var: Residual	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Num. groups elecid.i	290	290	290	290	290	290	290
Num. groups elecid.j	290	290	290	290	290	290	290
AIC	-218,962	-218,959	-218,948	-219,467	-219,123	-219,301	-229,142
BIC	-218,838	-218,834	-218,814	-219,333	-218,989	-219,167	-229,008
LL	109,494	109,492	109,488	109,748	109,576	109,665	114,585
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel models with non-hierarchical random intercepts for elections; decade FEs included but not shown; levels of significance: *** p < 0.001, ** p < 0.01, * p < 0.05.

Appendix D: Alternative model specifications

Below, we report alternative approaches for modeling the complex structure of the data. In addition, we test weighted vote gains/losses for the interaction models.

Alternative specifications of random and fixed effects structure

Gilardi and Füglister (2008, 425) suggest including three random effects (RE) in dyadic settings, one for each state plus time. Likewise, Schmidt-Catran and Fairbrother (2016, 34) argued that "[t]he first and general rule is to include REs at all levels at which there are FEs". The equivalent for states when analyzing parties would be the party. As an alternative approach one may include fixed effects (FE) to control for unobserved unit peculiarities. Regarding fixed effects though, Plümper et al. (2005, 331) warned that "unit dummies *completely absorb* differences in the *level* of independent variables across units" [emphasis in original]. Meyer (2013, 225–28) summarized the complex structure of observations at the party level and their hierarchical nesting in countries and elections. Accordingly, he suggests specifying different models, once with parties and once with elections as an RE. Considering these facts, we re-estimate the main models with a different random intercept and/or fixed effect structure. In particular, we test:

1. two non-hierarchical REs for the sender and receiver (i.e. parties);
2. two hierarchical REs for the sender and receiver, each nested in their respective country;
3. two hierarchical REs for the sender and receiver, each nested in their respective election;
4. a fixed effects specification for parties.

The results hardly change and support our substantive conclusions that diffusion takes place foremost in the regional context. The effect of being in the same EP faction is slightly stronger while diffusion among governments and within transnational party organizations is slightly less pronounced when controlling for "party peculiarities". Regarding the ordering, *Transnational party organizations* and *EP factions* switch places at times. Both are more important than diffusion from or among government parties, though. Thus, our finding of ties differing in their effect and their ordering is quite robust to alternative model specifications.

Table 6: Random effects for parties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
"Recycling"	0.405^{***}	0.405^{***}	0.404^{***}	0.398^{***}	0.400^{***}	0.358^{***}	0.351^{***}
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Competitors	0.227^{***}	0.227^{***}	0.227^{***}	0.228^{***}	0.228^{***}	0.181^{***}	0.181^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		-0.001^{*}					0.004^{***}
		(0.001)					(0.001)
Among governments			0.009^{***}				0.009^{***}
			(0.001)				(0.001)
EP factions				0.018^{***}			0.014^{***}
				(0.001)			(0.001)
Transnational party					0.012^{***}		0.003^{*}
organizations					(0.001)		(0.001)
Family of nations						0.062^{***}	0.062^{***}
						(0.001)	(0.001)
Intercept	0.268^{***}	0.268^{***}	0.266^{***}	0.266^{***}	0.267^{***}	0.253^{***}	0.249^{***}
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: party.i (Intercept)	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Var: party.j (Intercept)	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Var: Residual	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Num. groups party.i	162	162	162	162	162	162	162
Num. groups party.j	162	162	162	162	162	162	162
AIC	-219,620	-219,609	-219,738	-219,815	-219,734	-229,571	-229,820
BIC	-219,514	-219,494	-219,624	-219,700	-219,620	-229,456	-229,666
LL	109,821	109,816	109,881	109,919	109,879	114,797	114,926
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel model with non-hierarchical random intercepts for parties; decade FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 7: Random effects for parties nested in countries

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
"Recycling"	0.405***	0.405***	0.404***	0.398***	0.400***	0.358***	0.351***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Competitors	0.227***	0.227***	0.227***	0.228***	0.228***	0.181***	0.181***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		-0.001					0.004***
		(0.001)					(0.001)
Among governments			0.009***				0.009***
			(0.001)				(0.001)
EP factions				0.018***			0.014***
				(0.001)			(0.001)
Transnational party					0.012***		0.003*
organizations					(0.001)		(0.001)
Family of nations						0.062***	0.062***
						(0.001)	(0.001)
Intercept	0.277***	0.277***	0.276***	0.276***	0.276***	0.262***	0.259***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: party.i:iso.i (Intercept)	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Var: party.j:iso.j (Intercept)	0.002	0.002	0.002	0.002	0.002	0.001	0.001
Var: iso.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.001	0.001
Var: iso.j (Intercept)	0.002	0.002	0.002	0.002	0.002	0.001	0.001
Var: Residual	0.007	0.007	0.007	0.007	0.007	0.007	0.007
Num. groups party.i:iso.i	162	162	162	162	162	162	162
Num. groups party.j:iso.j	162	162	162	162	162	162	162
Num. groups iso.i	19	19	19	19	19	19	19
Num. groups iso.j	19	19	19	19	19	19	19
AIC	-219,753	-219,742	-219,872	-219,948	-219,869	-229,696	-229,946
BIC	-219,629	-219,608	-219,738	-219,814	-219,735	-229,563	-229,774
LL	109,890	109,885	109,950	109,988	109,949	114,862	114,991
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel model with non-hierarchical random intercepts for parties nested in countries; decade FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 8: Random effects for parties nested in elections

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
"Recycling"	0.406***	0.406***	0.406***	0.402***	0.403***	0.358***	0.355***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Competitors	0.228***	0.228***	0.228***	0.228***	0.228***	0.180***	0.181***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		-0.006***					0.008**
		(0.001)					(0.003)
Among governments			0.007***				0.012***
			(0.001)				(0.003)
EP factions				0.010***			0.005***
				(0.001)			(0.001)
Transnational party					0.010***		0.004***
organizations					(0.001)		(0.001)
Family of nations						0.064***	0.064***
						(0.000)	(0.000)
Intercept	0.265***	0.267***	0.264***	0.265***	0.264***	0.250***	0.246***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: party.i:elecid.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: party.j:elecid.j (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: elecid.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: elecid.j (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: Residual	0.004	0.004	0.004	0.004	0.004	0.003	0.003
Num. groups party.i:elecid.i	1314	1314	1314	1314	1314	1314	1314
Num. groups party.j:elecid.j	1313	1313	1313	1313	1313	1313	1313
Num. groups elecid.i	290	290	290	290	290	290	290
Num. groups elecid.j	290	290	290	290	290	290	290
AIC	-272,725	-272,768	-272,785	-272,829	-272,853	-291,804	-291,886
BIC	-272,600	-272,634	-272,652	-272,695	-272,719	-291,670	-291,714
LL	136,375	136,398	136,407	136,429	136,441	145,916	145,961
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel model with non-hierarchical random intercepts for parties nested in elections; decade FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 9: Fixed effects for parties

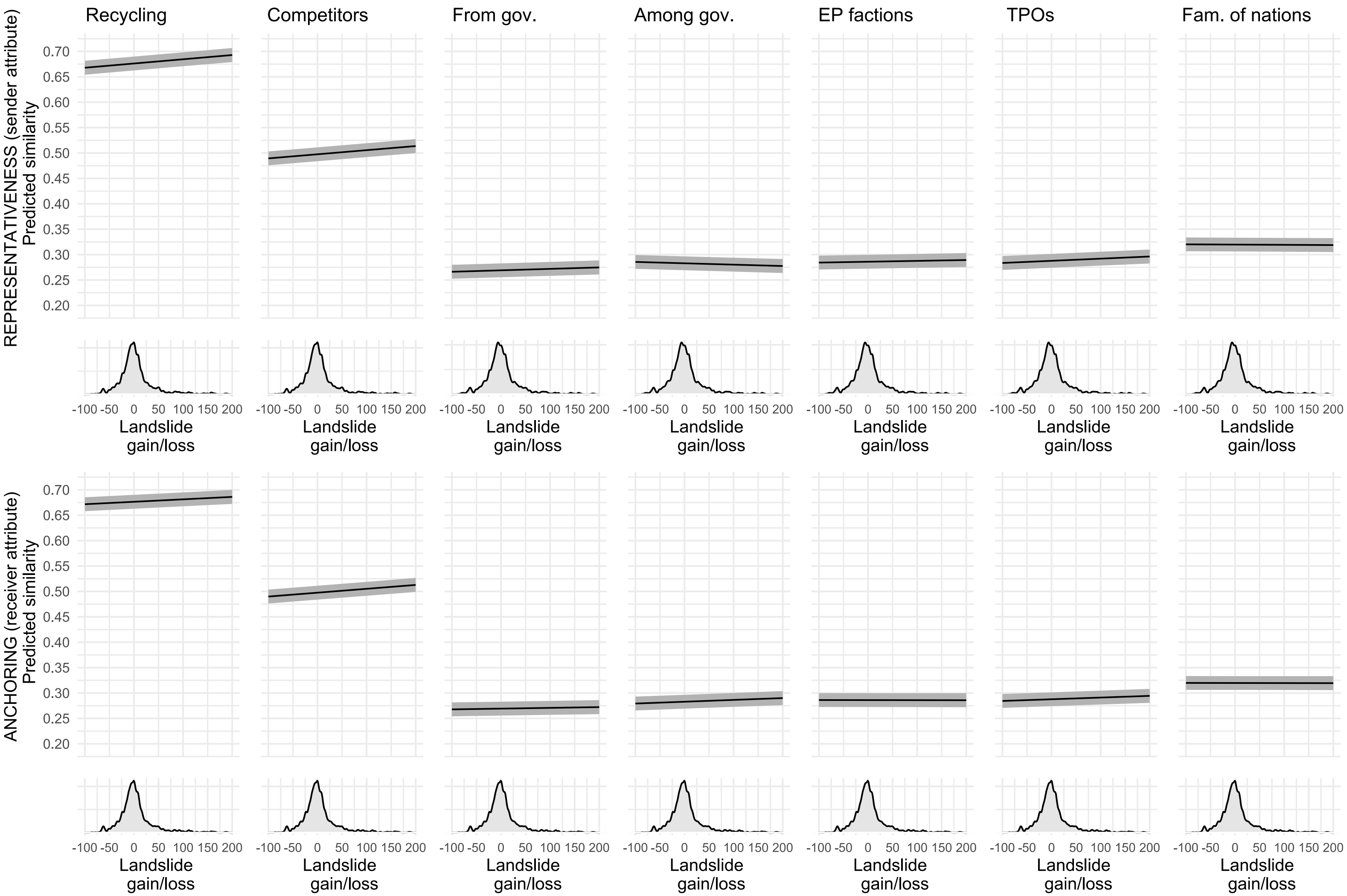
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
"Recycling"	0.405***	0.405***	0.404***	0.398***	0.400***	0.358***	0.351***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Competitors	0.227***	0.227***	0.227***	0.228***	0.228***	0.181***	0.181***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		-0.001*					0.004***
		(0.001)					(0.001)
Among governments			0.008***				0.009***
			(0.001)				(0.001)
EP factions				0.018***			0.014***
				(0.001)			(0.001)
Transnational party					0.012***		0.003*
organizations					(0.001)		(0.001)
Family of nations						0.062***	0.062***
						(0.001)	(0.001)
Intercept	0.192***	0.192***	0.194***	0.190***	0.191***	0.181***	0.180***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Party_i$ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$Party_j$ FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.552	0.552	0.553	0.553	0.553	0.592	0.594
Adj. R^2	0.551	0.551	0.551	0.551	0.551	0.591	0.592
RMSE	0.085	0.085	0.085	0.085	0.085	0.081	0.081
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Linear model with fixed effects for sender and receiver party; decade and party FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Alternative specification of sender and receiver attributes

Instead of unweighted vote gain/loss (in percentage points), we re-run the interaction models with "landslide gain/loss". Landslide relates a gain or loss to a party's size, i.e. it captures relative gains/losses as percent of *i* or *j*'s past (or previous) vote share. This operationalization downplays minor changes and emphasizes remarkable instances of success or loss. The rationale is the critique that a gain or loss of five percentage points "means" different things for large parties (say, with a vote share of >40 per cent) or smaller ones (<10 per cent). While the effects are slightly less pronounced than for the "simple" gain/loss, in principle the results point in the same direction supporting our substantive interpretations.

Figure 3: Conditional effects of sender and receiver attribute using "landslide gain/loss"



Notes: Predictions with 90% CIs adjusting for all other covariates and assuming RE=0; the bottom graphs show the kernel density of observed data for *Landslide gains/losses*.

Appendix E: Text similarity and left-right scores

So far, studies analyzing party policy diffusion have assumed that party positions accurately capture *what* diffuses. However, left-right scores in particular have been challenged on several grounds (cf. for example the 2007 special issue of *Electoral Studies* [Marks et al. 2007], Benoit et al. 2009; Hooghe et al. 2010; Laver 2001; Volkens et al. 2013). Our suggestion to overcome some of these issues and move closer to the content of diffusion is to treat words as data (Laver et al. 2003). At least since the 2000s, quantitative text analysis has matured and such approaches are becoming mainstream in comparative policy analysis (Gilardi and Wüest 2018). Arguing that diffusion is a process best traced on the level where ideas materialize – the text – cosine similarity of texts is an appropriate measure for this purpose. Regarding the thus far most prevalent operationalization of party policy diffusion via the Manifesto Group's RILE index (Budge and Klingemann 2001), the question emerges whether and to what extent these measures are interchangeable?

We first examine the bivariate relationship between the two quantities. We then recalculate our models including absolute differences of the RILE. It has been argued that text similarity captures everything else but the diffusion of party policies. We agree that text similarity indeed captures other aspects like rhetoric, style, or pledges as well. Being interested in the "grand picture" and the impact of ties and sender/receiver attributes in shaping diffusion, our analysis still shows the potential of text-as-data approaches and is likely underestimating diffusion effects.

While we do not share this assumption, assume for now that RILE scores better capture the similarities of meaningful manifesto content. Following this line of argument, parties that are close on the left-right dimension (i.e. have a very low absolute difference in their RILE positions) should have a high text similarity. Reversely, parties that are far away on the RILE should have a very low text similarity. If it holds, the absolute difference included as a "control" should determine text similarity to such an extent that it absorbs other effects in the models. Empirically, this is not true.

Figure 4 shows the scatter plot of both variables and Table 10 reports the results when "controlling" for party positions. As expected, there is a negative, but weak correlation (Pearson's r = -0.16). Likewise, our models hardly change in terms of effect size or direction regarding the impact of ties. This indicates that RILE scores capture some aspects of text similarity but are *not* interchangeable. Furthermore, they may misleadingly signal similarities when there are none, explain text similarities only to a slight extent, and are not well equipped for digging deeper into the content of diffusion. We therefore conclude that textual similarity is a better measure for analyzing diffusion among parties.

Figure 4: Scatter plot of cosine similarity and abs. distance on the left-right dimension (RILE)

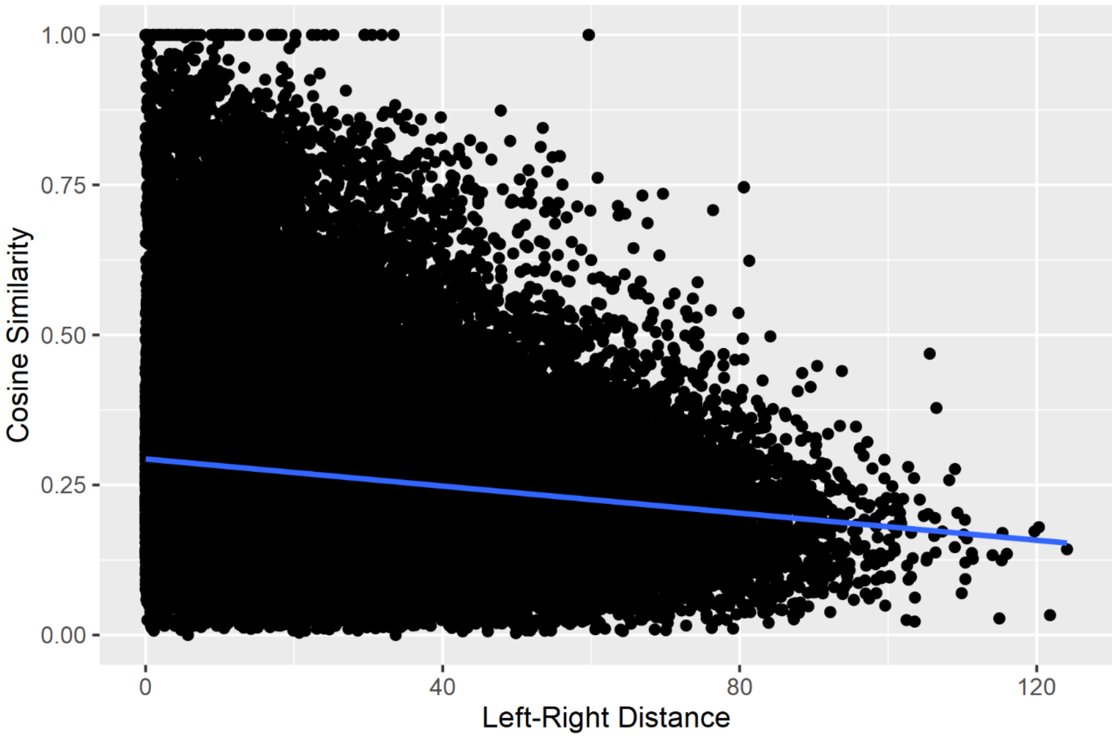


Table 10: Regression models incl. RILE distances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Abs. Δ RILE	-0.047***	-0.047***	-0.046***	-0.046***	-0.045***	-0.047***	-0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
"Recycling"	0.401***	0.400***	0.399***	0.395***	0.394***	0.353***	0.347***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Competitors	0.227***	0.227***	0.227***	0.228***	0.228***	0.180***	0.180***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		-0.001					0.005***
		(0.001)					(0.001)
Among governments			0.015***				0.015***
			(0.001)				(0.001)
EP factions				0.014***			0.003
				(0.001)			(0.001)
Transnational party					0.017***		0.011***
organizations					(0.001)		(0.001)
Family of nations						0.064***	0.063***
						(0.001)	(0.001)
Intercept	0.282***	0.282***	0.279***	0.281***	0.280***	0.267***	0.261***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Decade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: elecid.i (Intercept)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Var: elecid.j (Intercept)	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Var: Residual	0.007	0.007	0.007	0.007	0.007	0.006	0.006
Num. groups: elecid.i	290	290	290	290	290	290	290
Num. groups: elecid.j	290	290	290	290	290	290	290
AIC	-219,953	-219,941	-220,395	-220,064	-220,206	-230,208	-230,807
BIC	-219,838	-219,817	-220,271	-219,940	-220,081	-230,083	-230,645
LL	109,989	109,984	110,210	110,045	110,116	115,117	115,421
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Abs. Δ RILE divided by 100 for optical reasons; multilevel models with non-hierarchical random intercepts for elections; decade FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Appendix F: Comparing Jaccard and cosine similarity

We argue that the outcome of diffusion processes materializes as text similarity. As a first step to overcome left-right scores, we therefore measure how close two texts are to each other. Different methods have been developed to represent texts as vectors and to further process this numerical correspondence (Bär et al. 2015). Applying a bag-of-words approach, we consider the overlap of the vocabulary and the frequency of words ("emphasis"). Cosine similarity is an appropriate measure for this task.

The bag-of-words approach transfers all documents into vectors (Grimmer and Stewart 2013). Cosine similarity then quantifies the angle between two vectors indicating to what extent two documents A and B point in a similar direction:

$$\cos(x) = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

Cosine similarity takes the unique set of words into account, but – unlike Jaccard similarity – it also considers the length of the vectors. Because it takes both aspects into account, a change in text similarity, however, cannot be attributed to one of these two factors alone. As an alternative, one may apply Jaccard similarity for measuring the similarity of two documents. It ignores term frequencies and solely looks at two sets of unique words and the size of their union.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Contrary to cosine similarity, the Jaccard coefficient is more intuitive regarding its interpretation at the document or vocabulary level: A Jaccard coefficient of 0.10 means that 10 percent of the vocabulary of both documents is shared.

An issue with applying Jaccard similarity to election programs is that diffusion of ideas occurs not only by adopting previously unused terms, but also by aligning the frequency of word usage. Jaccard similarity has a significant blind spot here. Moreover, as the lengths of texts increase, the probability that it contains a common term increases; the consequence is that longer texts often show a greater overlap of the vocabulary. On the other hand, Jaccard similarity decreases when comparing a long with a short text as the likelihood for a larger number of total unique words increases for longer texts thereby increasing the denominator while hardly affecting the nominator. Taken together, the properties of the Jaccard coefficient usually result in lower text similarity.

In Figure 5 below we depict three scenarios for comparing the relationship of Jaccard and cosine similarity: (1) a party *adds* previously unused terms from a remote document, (2) a party *replaces* some of its own words with terms from the other document, or (3) a party *increases its emphasis* of one word. In the first and second case the vocabulary changes but not the frequency of words, and vice versa for the third example.

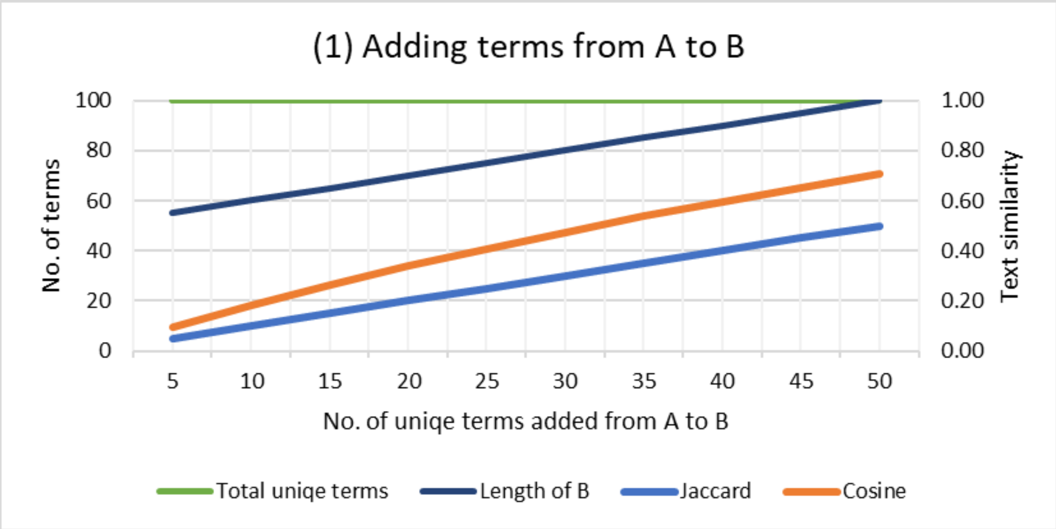
(1) For scenario 1 and 2 imagine two manifestos A ("template from sender") and B ("first draft of new manifesto"). Each consists of 50 unique terms which, in turn, are mentioned only once (i.e. 100 in total with no overlap). A party now adds previously unused terms from A to its draft. The total number of unique words does not change, but the length of B increases. For every word that is added from A to B, Jaccard similarity increases by 1 percent. Cosine similarity increases at a faster rate as it takes the increasing length and overlap into account.

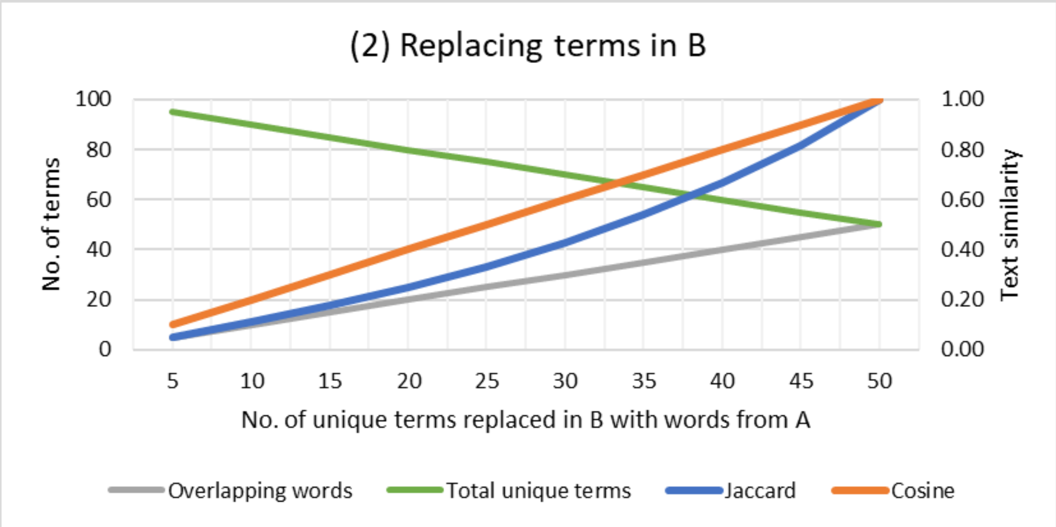
(2) In the second case, a party adapts previously unused terms from the remote document; this time, however, replacing some of their own terms. The length of B does not change and the overlap increases. Consequently, the total number of unique terms decreases thereby affecting the denominator. Under these circumstances, cosine similarity increases linearly while Jaccard increases exponentially.

(3) For the third scenario imagine two documents C and D. Both consist of 10 unique terms, each mentioned five times. In addition, both have one word in common which is mentioned 25 times in C but only once in D. Now, a party aligns its emphasis by mentioning this word more often. The length of D increases until it is on par with C. As the number of unique terms (21 in total) does not change, Jaccard similarity remains constant. Cosine similarity, however, is able to detect the "convergence".

In reality, parties will adjust both the vocabulary and the frequency of words. For this reason, cosine similarity is better suited to measure the outcome of diffusion. Still, in Table 11 and Figure 6 we report the results for our analysis using Jaccard similarity as the dependent variable. As expected, the effect strengths are lower compared to cosine similarity. However, the substantive conclusions remain, with ties differing in their strength and order. Again, diffusion in the regional context is more important than diffusion within EP factions and transnational party organizations, and less relevant from and among government parties. The interactions in turn resemble the results for cosine similarity pointing in the same direction.

Figure 5: Comparing Jaccard and cosine similarity





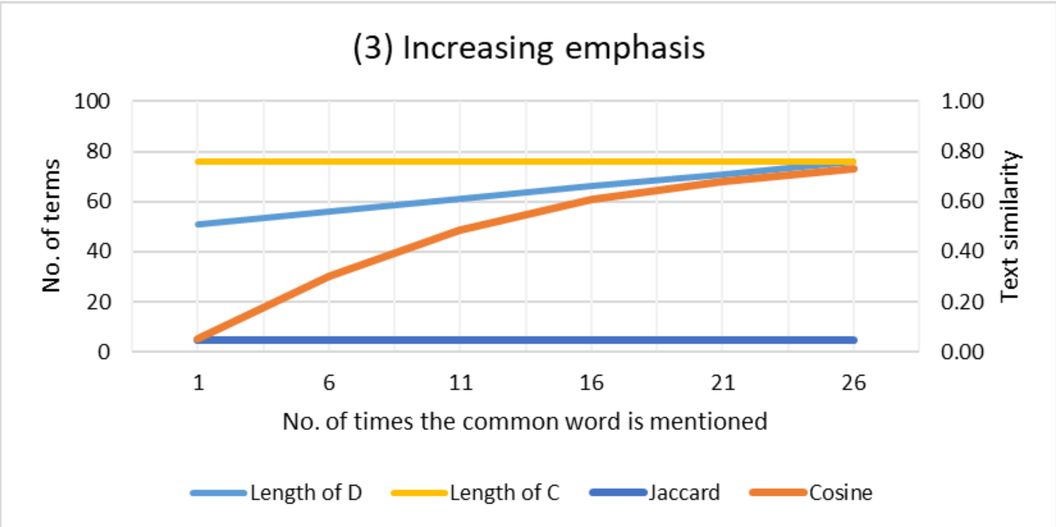
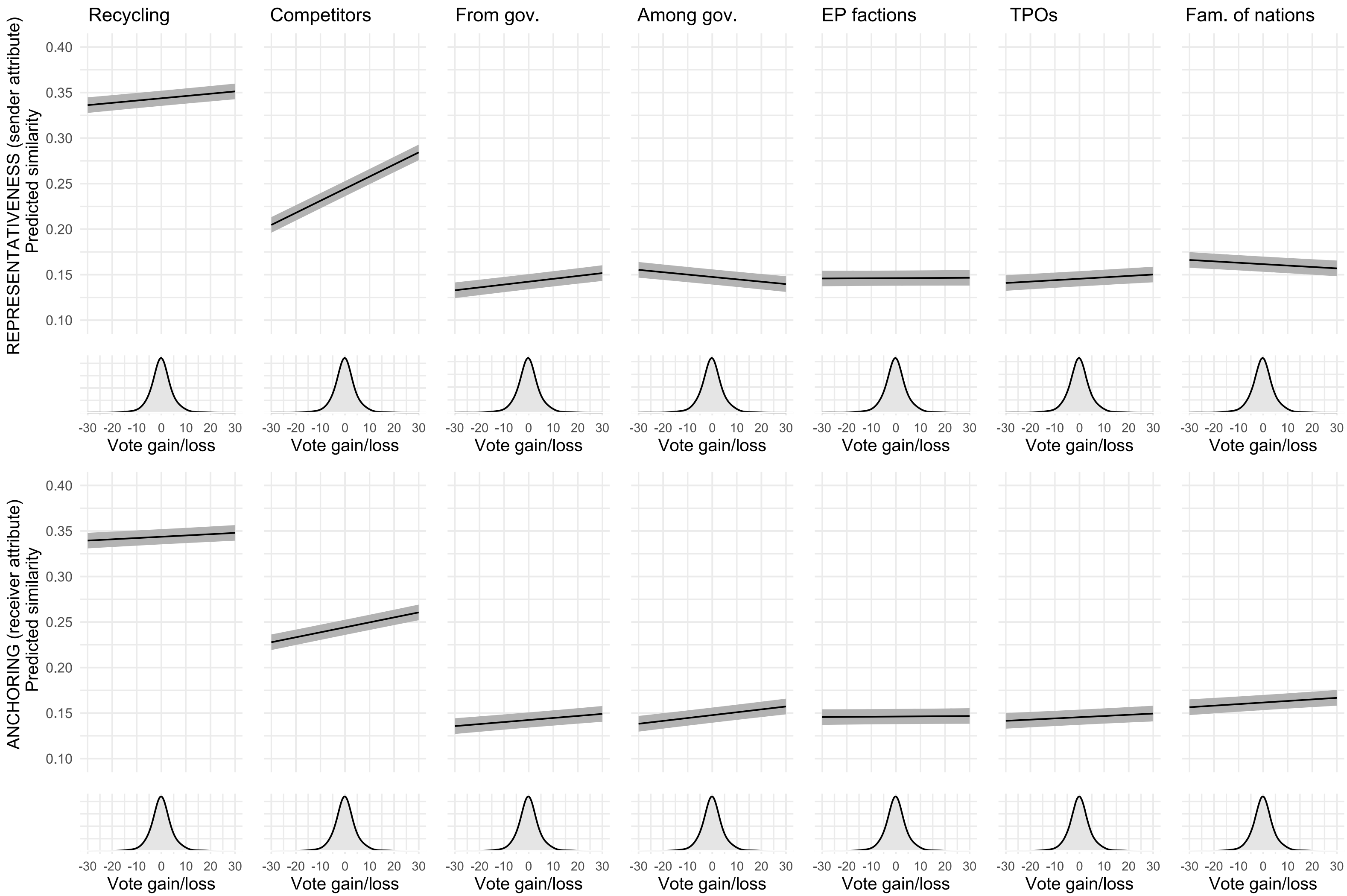


Table 11: Regression model for each tie with Jaccard similarities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
"Recycling"	0.202***	0.202***	0.201***	0.200***	0.200***	0.183***	0.181***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Competitors	0.102***	0.102***	0.102***	0.102***	0.102***	0.083***	0.083***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
From governments		0.001					0.004***
		(0.000)					(0.000)
Among governments			0.007***				0.008***
			(0.000)				(0.000)
EP factions				0.004***			0.002*
				(0.001)			(0.001)
Transnational party					0.004***		0.001
organizations					(0.001)		(0.001)
Family of nations						0.025***	0.025***
						(0.000)	(0.000)
Intercept	0.142***	0.142***	0.141***	0.142***	0.142***	0.136***	0.134***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Decade Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Random Parts							
Var: elecid.i (Intercept)	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Var: elecid.j (Intercept)	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Var: Residual	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Num. groups: elecid.i	290	290	290	290	290	290	290
Num. groups: elecid.j	290	290	290	290	290	290	290
AIC	-313,305	-313,291	-313,553	-313,320	-313,322	-317,048	-317,313
BIC	-313,199	-313,176	-313,438	-313,205	-313,207	-316,933	-317,159
LL	156,663	156,657	156,788	156,672	156,673	158,536	158,672
Obs.	105,575	105,575	105,575	105,575	105,575	105,575	105,575

Note: Multilevel models with non-hierarchical random intercepts for elections; decade FEs included but not shown; levels of significance: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Figure 6: Conditional effects of sender and receiver attributes on Jaccard similarity



Notes: Predictions with 90% CIs adjusting for all other covariates and assuming RE=0; the bottom graphs show the kernel density of observed data for *Vote gains/losses*.

References

Bär D, Zesch T and Gurevych I. 2015. Composing Measures for Computing Text Similarity. Technical Report TUD-CS-2015-0017.

Benoit K *et al.* (2018). quanteda: An R package for the quantitative analysis of textual data. *The Journal of Open Source Software* **3** (30), 774.

Benoit K, Laver M and Mikhaylov S (2009). Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions. *American Journal of Political Science* **53** (2), 495–513.

Böhmelt T *et al.* (2016). Party Policy Diffusion. *American Political Science Review* **110** (2)397-410.

Böhmelt T *et al.* (2017). Why Dominant Governing Parties Are Cross-Nationally Influential. *International Studies Quarterly* **61** (4), 749–759.

Budge I and Klingemann H-D (2001). Finally! Comparative Over-Time Mapping of Party Policy Movement. In Budge I, Klingemann H-D, Volkens A, Bara JL and Tanenbaum E (eds), *Mapping Policy Preferences:* Estimates for Parties, Electors, and Governments, 1945-1998. Oxford: Oxford University Press, pp. 19–50.

Castles FG (1998). *Comparative Public Policy: Patterns of Post-war Transformation.* Cheltenham: Elgar.

Gilardi F and Füglister K (2008). Empirical Modeling of Policy Diffusion in Federal States: The Dyadic Approach. *Swiss Political Science Review* **14** (3), 413–450.

Gilardi F and Wüest B (2018). *Text-as-Data Methods for Comparative Policy Analysis.* University of Zurich.

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

Hooghe L *et al.* (2010). Reliability and Validity of Measuring Party Positions: The Chapel Hill Expert Surveys of 2002 and 2006. *European Journal of Political Research* **49** (5), 687–703.

Huang A (2008). Similarity Measures for Text Document Clustering, Christchurch, New Zealand.

Laver M (2001). How should we estimate the policy positions of political actors? In Laver M (ed), *Estimating the Policy Position of Political Actors*. London: Routledge, pp. 239–244.

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

Le QV and Schuster M (2016). A Neural Network for Machine Translation, at Production Scale. Available from https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html (accessed 13 February 2019).

Lotz S and van Rensburg A (2014). Translation Technology Explored: Has a Three-Year Maturation Period Done Google Translate Any Good? *Stellenbosch Papers in Linguistics Plus* **43**, 235–259.

Lucas C *et al.* (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis* **23** (2), 254–277.

Marks G *et al.* (2007). *Special Symposium: Comparing Measures of Party Positioning: Expert, Manifesto, and Survey Data.* Electoral Studies 26 (1).

Meyer TM (2013). *Constraints on Party Policy Change.* Colchester: ECPR Press.

Plümper T, Troeger VE and Manow P (2005). Panel Data Analysis in Comparative Politics: Linking Method to Theory. *European Journal of Political Research* **44** (2), 327–354.

Reber U (2018). Overcoming Language Barriers: Assessing the Potential of Machine Translation and Topic Modeling for the Comparative Analysis of Multilingual Text Corpora. *Communication Methods and Measures* **3** (4), 1–24.

Schmidt-Catran AW and Fairbrother M (2016). The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right. *European Sociological Review* **32** (1), 23–38.

Volkens, Andrea, Judith Bara, Ian Budge, Michael D. McDonald, and Hans-Dieter Klingemann, eds. (2013). *Mapping Policy Preferences from Texts: Statistical Solutions for Manifesto Analysts.* Oxford: Oxford University Press.

Vries E de, Schoonvelde M and Schumacher G (2018). No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications. *Political Analysis* **26** (4), 417–430.

Welbers K, van Atteveldt W and Benoit K (2017). Text Analysis in R. *Communication Methods and Measures* **11** (4), 245–265.