

Overcoming Language Barriers: Assessing the Potential of Machine Translation and Topic Modeling for the Comparative Analysis of Multilingual Text Corpora

Ueli Reber

University of Bern, Switzerland

Institute of Communication and Media Studies, Fabrikstrasse 8, CH-3012 Bern

ueli.reber@ikmb.unibe.ch

Abstract: This study assesses the potential of topic models coupled with machine translation for comparative communication research across language barriers. From a methodological point of view, the robustness of a combined approach is examined. For this purpose the results of different machine translation services (Google Translate vs. DeepL) as well as methods (full-text vs. term-by-term) are compared. From a substantive point of view, the integratability of the approach into comparative study designs is tested. For this, the online discourses about climate change in Germany, the United Kingdom, and the United States are compared. First, the results show that the approach is relatively robust and second, that integration in comparative study designs is not a problem. It is concluded that this as well as the relatively moderate costs in terms of time and money makes the strategy to couple topic models with machine translation a valuable addition to the toolbox of comparative communication researchers.

Keywords: machine translation, topic model, comparative research, multilingual analysis, climate change, online discourse

Funding: This publication was created in the context of the Research Unit “Political Communication in the Online World” (FOR 1381), Subproject 07, which is funded by the German Research Foundation (DFG, project number 155794648). The subproject is also funded by the Swiss National Science Foundation (SNSF, project number 100017E-154100).

Introduction

In communication research, comparing instances of public communication has become a popular approach to gain a better understanding of publics and discourses. This is because the comparative perspective allows us to draw conclusions beyond the individual case, either by testing theories in diverse settings or by evaluating the scope and significance of certain phenomena in different contexts (Esser & Hanitzsch, 2012). Even more so in times of digitalization, in which national borders are becoming increasingly permeable for communication flows, the comparative perspective is important to answer the question whether a phenomenon is common across the globe or distinctive to a certain country (Esser, 2013; Livingstone, 2012).

However, comparing instances of public communication is theoretically as well as methodologically challenging (cf. Livingstone, 2003). This is especially true if the compared cases happen to be in different languages. For such comparisons, not only must instruments be developed that work for all studied cases, but also must there be people with the right language skills to collect and analyse the data. This makes analyses of multilingual text corpora particularly complex, labor-intensive, and costly. It is therefore of great interest that now well-established automated content analysis methods can also be used for comparisons of public communication across multiple languages.

One such method, which is now a standard tool in communication research (Boumans & Trilling, 2016; Grimmer & Stewart, 2013), are so-called bag-of-words topic models. These are generative models which allow the identification of underlying thematic structures (i.e., topics) even in large amounts of text (Jacobi, van Atteveldt & Welbers, 2016). They are called bag-of-words models because the order of the words in a document (i.e., its syntactic structure) is ignored in the modeling process. However, they are not designed for the analysis of multilingual corpora as the resulting topics directly depend on the vocabulary used in the

documents (cf. Blei, Ng & Jordan, 2003). Since languages have different vocabularies, topic models separate topics by language. Standard topic models are thus blind for thematic structures which cross language boundaries.

To solve this “confusion of tongues,” there are two possible ways: either compute a topic model on the multilingual corpus and match the topics after or translate the corpus into a common language first and then compute the model. The first approach is challenging because additional internal or external information is needed to bring the topics together (e.g., Wikipedia entries on the same topic in different languages, as used in the Polylingual Topic Model by Mimno, Wallach, Naradowsky, Smith, and McCallum, 2009). The second approach is easier, given that a low-cost and reliable translation of the corpus is possible. Thanks to improvements in cheap machine-translation services such as Google Translate (Lotz & van Rensburg, 2014), this seems to be a feasible option. In fact, it has recently been shown that machine translation and topic modeling can be combined to study public communication (Lucas et al., 2015; de Vries, Schoonevelde & Schumacher, 2018).

This article continues along this path by assessing the potential of such a combined approach for comparative communication research. From a methodological point of view, the robustness of the approach is examined by determining whether the choice of translation service and method matters when it is used for analytical purposes. From a substantive point of view, it is examined whether topic modeling coupled with machine translation can be linked to existing theories and work from the field of comparative communication research in such a way that it is a valuable add-on to the toolbox. For both, the online discourses about climate change in Germany, the United Kingdom and the United States of America are used as test cases. As a global phenomenon, climate change is a widely studied topic in the field of comparative communication research. However, the focus has been on offline media so far. A multilingual comparison of online discourses is uncharted territory and thus an ideal research

subject to assess the potential of topic modeling in combination with machine translation for comparative analyses.

Study 1: Robustness

De Vries, Schoonvelde, and Schumacher (2018) have shown that the translations of professional human translators (“gold standard”) and Google Translate are quite similar when documents are considered a “bag of words” (i.e., ignoring the syntactic structure) and thus concluded that machine translation services are indeed useful tools when it comes to the analysis of multilingual text corpora with topic models.

But how much do translation results of different machine translation providers differ? Or to put it another way: Does it matter which translation provider is used for the translation of a multilingual text corpus when topic models are used for its analysis? To answer this question, this study compares two full-text translations of the same corpus—one by Google Translate and one by DeepL. This is the current top dog (Google Translate) compared to the rising star (DeepL) in the field of online machine translation providers. According to their own surveys, DeepL currently provides the best translations of all online translation services available (DeepL, 2017). However, it remains to be seen whether the better results of DeepL—in terms of readability—are also of importance in bag-of-word analyses or if Google Translate does the job equally well.

In their assessment, de Vries et al. (2018) focused on translations of whole documents. However, since most machine-translation services charge fees based on the number of translated characters, it would be significantly more cost-effective (as well as time-saving) if every word in a corpus has to be translated only once. The good thing about topic models is that they perceive documents just as vectors containing the count of each word within the document, ignoring the order in which they appear (i.e., bag of words). For the modeling process, all documents of a corpus are bound together in a so-called document-term matrix

(DTM), containing the frequency of each word (term) in each document. The translation of the unique terms of a DTM (i.e., its vocabulary) would thus be an extremely efficient way to obtain a monolingual corpus, as every word must be translated only once instead of multiple times for every document. One potential problem with such a term-by-term translation, however, is that the absence of the syntactic structure may affect the outcome of the translation. The question thus is whether this shortcut is a valid alternative to the more expensive full-text translation. Or more precisely: Does it matter which translation method is used for topic model analyses?

Thus far, only Lucas et al. (2015) compared a DTM translation to a full-text translation of the same corpus in combination with topic modeling. After translating Chinese and Arabic tweets into English, they computed topic models on both corpora and assessed their similarity by comparing the content of the topics (i.e., their word probabilities). They concluded that “two investigators using different translation methods might have reached similar substantive conclusions” (Lucas et al., 2015, p. 274). This result is checked here by systematically comparing the term-by-term translation of a corpus (DTM translation) to the full-text translations of the same corpus mentioned above. Google Translate was chosen for the DTM translation. There are two reasons for this choice: First, Google Translate was also the choice of Lucals et al. (2015) for their DTM translation. This allows the translation service to be kept constant. Second, DeepL restricts the use of their API in a way that a DTM translation does not save time (and money) compared to a full-text translation.¹

Study 2: Integrability

The number of socioscientific studies that have explicitly used machine translation is still limited. Thus far, most studies have had a methodological focus, as they either proposed and evaluated concrete analytical procedures that build on machine-translated texts (Agarwal,

1 This changed with the introduction of the DeepL Pro plan. For this study, however, an earlier and more restricted version of the API was used. Due to the restrictions (i.e., limited number of translation requests per time), the translation of the DTM would have taken considerably longer than the full-text translation.

Xie, Vovsha, Rambow, & Passonneau, 2011; Balahur & Turchi, 2014; Pennings, 2011), evaluated the quality of machine-translation services (e.g., Hampshire & Porta Salvia, 2010; Lotz & van Rensburg, 2014), or assessed their usefulness for analyses with text models (de Vries et al., 2018; Lucas et al., 2015). Studies that have used machine translation to examine substantive questions are especially rare. Benoit, Schwarz, and Traber (2012) translated parts of their corpus of parliamentary speeches for an analysis of legislators' policy preferences and Zhou, Cristea and Roberts (2015) used Google Translate for a sentiment analysis of war-related Wikipedia articles. In any case, the possible value of machine translation and topic modeling for comparative communication research has hardly been shown (the study by Lucas et al. (2015) is the notable exception). For this reason, a substantive question is examined here. Specifically, it is examined whether actors in Germany, the United Kingdom, and the United States debate the same aspects of climate change on their websites and whether they accord them similar relevance.

Since climate change is a global problem that might only be solved globally, it is important to know whether the public discourse about it is rather national or transnational in nature. To assess that, issue networks are studied here. That are networks of public communication, consisting of the websites as well as the hyperlinks connecting the contents of these websites (Marres & Rogers, 2005). Bennett, Lang and Segerberg (2015) argue that such issue networks are full-fledged issue publics, as they are open to connection and contestation among all kinds of actors. One way to measure the transnationality of such issue networks is through the analysis of hyperlink connections (e.g. Bennett et al., 2015). Another way is to analyze the contents of the websites. The second approach was chosen here, as *discourse convergence* has been identified as one of the main indicators for the transnationality of public discourses (Eder, 2000; Eder & Kantner, 2000; Kantner, 2004). At the core stands the question of “whether speakers in different national public spheres identify

the same issue as important, accord them similar relevance, and employ similar problem definitions” (Wessler, Peters, Brüggemann, Kleinen-von Königslöw, & Sifft, 2008, p. 11). From that point of view, a public discourse is transnational “if within an anonymous mass public the same issues are discussed at the same time under similar criteria of relevance” (Eder & Kantner, 2000, p. 315, translation by author). For the this study, it means that actors in Germany, the United Kingdom, and the United States have to discuss the same aspects of climate change on their websites and accord them similar relevance for it to be a transnationalized debate.

Although there is a rich and growing body of literature regarding online climate change communication (for an overview, see Schäfer, 2012), comparative studies are still the exception (e.g., Häussler, Adam, Schmid-Petri, & Reber, 2017; Jang & Hart, 2015). The content of the public debate about climate change has thus far only been compared in regard to Twitter (Kirilenko & Stepchenkova, 2014) and the mass media (e.g., Grundmann & Scott, 2014; Ivanova, 2017; Ivanova, Schmidt, & Schäfer, 2014; Schmidt, Ivanova, & Schäfer, 2013) across countries and languages.

This study therefore addresses three questions: First, how much do translation results of different machine translation providers differ? Second, is the translation of a DTM a valid alternative to the full-text translation of a corpus? Third, is the online climate change discourse the same in Germany, the United Kingdom, and the United States? All three questions aim to clarify whether the combination of machine translation and topic modeling are a valuable addition to the methodical toolbox of comparative communication researchers. In the following section, the data-collection procedure as well as the steps taken for the preprocessing and translation of the corpus is described. Then, the comparison of the translation providers as well as the translation methods is explained (study 1), and the procedure for the cross-country analysis of the climate change discourse is laid out (study 2).

The results of both studies are then presented. The paper concludes with a discussion of the results as well as the promises and limitations of the applied approach for comparative communication research.

Data and Method

Web Scraping and Building of the Corpus

The corpus for both studies—the comparisons of translation providers/methods and the comparison of the climate change discourse in Germany, the United Kingdom, and the United States—was collected in June 2014 in the course of a larger project conducted with collaborative partners (Adam et al., 2016). That project proceeded in four steps to gather all sorts of websites with an interest in climate change, which includes, but is not limited to, websites of civil society actors (such as NGOs, Blogs, universities, churches), the media, and governmental bodies.

A snowball-sampling strategy was used to harvest hyperlink-networks originating from carefully selected starting points in four countries (four climate advocates and four climate skeptics in Germany, the United Kingdom, and the United States; see the list in Appendix A). This was done with the help of the crawler software Issue Crawler (Rogers, 2013). The starting points were chosen based on literature reviews, expert interviews, and country-specific Google searches.

The crawled websites were then indexed according to keywords to ensure that they related to climate change, and they were downloaded if at least one keyword was mentioned once.²

2 Keywords: climate change, global warming, Klimawandel*, globale* Erwärmung

To obtain additional information about the actors included in this sample, the country of activity (national, transnational) as well as the name of the actor was coded manually by two coders using information found on the “about us” or a similar page.³

The final step was the extraction of plain text from the archived HTML documents. To do this, all HTML markups were deleted using the Java-based content extraction library Apache Tika. Then, meaningless terms used on most Web pages (e.g., navigation elements, copyright information) were filtered out by deleting all sentences containing “regular expressions” (i.e., character sequences defining search patterns) from a blacklist. The remaining documents were then categorized according to their language (German or English) by a language-detection algorithm and marked as duplicates if their similarity, defined by the Jaccard index on their word set, was above a threshold of 0.95.

For the analyses conducted in this study, the corpus was reduced by removing all duplicates, all websites from non-German, non-British, and non-U.S. American actors, and all websites not written in German or English. For June 2014, this resulted in a total of 875 unique Web pages published by 95 German actors, 2,172 Web pages published by 181 British actors, and 3,896 Web pages published by 539 U.S. American actors. From the total of 6,843 Web pages, 633 were written in German and 6,310 in English because some German actors also publish in English. For the assessment of the translation providers/methods, only the 633 Web pages written in German were considered.

Translation and Preprocessing

The translations process of the 633 German full-texts was straightforward: The documents were translated via Google Translate’s and DeepL’s API. However, due to length restrictions by the DeepL API, the texts were segmented into sentences, then translated, and finally put back together. The same was done for Google Translate to avoid a bias. All the

³ The intercoder reliability was measured by comparing the coders’ classification with a master coding. Krippendorff’s alpha was .93 for the country of activity (distinguishing 199 categories). The intercoder reliability was thus satisfactory.

textual data manipulation as well as the translation itself was done in *R* (R Core Team, 2017). For translations with Google Translate, the *translateR* package (Lucas & Tingley, 2014) was used. The package provides easy out-of-*R* access to the Google Translate API. A similar function was built to translate the documents with DeepL via their API.⁴

The process of the term-by-term translation was less straightforward, as the corpus had to be converted to a DTM first, requiring the documents to be preprocessed. The problem with most preprocessing steps is that they are not language-independent. This is true for common procedures like stop word removal, stemming, decompounding, or lemmatization. An attempt was made to include as much grammatical information as possible in the DTM for the translation. Thus, as little preprocessing as possible was done before the translation. No decompounding or stemming procedures were applied, and all the letters were kept as they were (uppercase or lowercase). However, all 633 German documents were split into individual words (tokenization), and then all punctuation marks (Unicode “Punctuation” [S] class), symbols (Unicode “Symbol” [S] class), and tokens consisting only of numbers were removed. Hyphens were not removed, as that might change the meaning of terms (e.g., “EU-Abgeordneter” [Member of the European Parliament]). From these tokens, a DTM was created. This step, as well as the rest of the preprocessing, was done in *R* (R Core Team, 2017) using the *Quanteda* package (Benoit et al., 2017). For the actual translation, the vocabulary (i.e., the unique terms) was extracted from the German DTM, translated with Google Translate using the *translateR* package (Lucas & Tingley, 2014) and placed back into the DTM as English terms.

A specific feature of the German language is the concatenation of multiple words into a single word. Because such compounds (e.g., “Klimawandel”) are less common in English, they were often translated as ngrams (e.g., “climate change”). To avoid systematic differences

⁴ The R code will be available on Github after the manuscript is accepted for publication. The URL will be inserted here.

in the vocabulary, all the newly created ngrams in the “German” DTM were split into unigrams (e.g., “climate” and “change”) before all further steps. In addition, duplicate terms were removed.

Apart from counting the words, the aim of preprocessing is to remove noise and to extract as much useful information as possible by “treating words with very similar properties identically and removing words that are unnecessary to our interpretation and our model” (Lucas et al., 2015, p. 257). Hence, the final preprocessing steps—which were conducted on all the translated DTMs for both studies—included the conversion of all letters to lowercase, the removal of terms with fewer than three characters, lemmatization, the removal of stop words, and relative pruning (for further reflection on preprocessing, see Maier et al., 2018). The preprocessing steps were conducted in this order. It is important to note that a different order would lead to different results (Denny & Spirling, 2017). The conversion to all lowercase was done for the sake of term unification. This was necessary because of words used, for example, at the beginning of a sentence. Stop word removal and the removal of terms with fewer than three characters (e.g., “to,” “on,” “dr,” “mp”) was done to remove terms that are extremely frequent or unspecific and thus not helpful as indicators for a document’s content (Salton, 1991). For stop word removal, a look-up list with the most common stop words (e.g., “from,” “after,” “the”) as well as stop words specific to websites (e.g., “click,” “login,” “comments”) was used. The list was put together specifically for this project. Lemmatization was done to convert inflected words to their base forms (e.g., “warming” to “warm”). Since the declination or conjugation of a word is usually not indicative of its meaning, lemmatization is used to combine words with the same meaning (Lucas et al., 2015). This reduces the dimension of the model input and generally improves the results (Jacobi et al., 2016). A comprehensive look-up list with base forms of inflected words was used for the lemmatization. Finally, relative pruning was done to remove

extremely rare as well as extremely frequent terms. All terms were removed that occurred in less than 0.5% or more than 99% of the documents, two common thresholds (e.g., Grimmer & Steward, 2013). If there is a theoretical interest in finding patterns of words that are used across documents, the removal of very infrequent terms is recommended because they do not contribute much information to the identification of document similarities (Denny & Spirling, 2017, p. 8). The removal of very frequent terms is advisable for the same reason as the removal of stop words—they are unspecific for a document’s content and therefore add no helpful information to the topic model.

The Structural Topic Model

For the analyses done here, four topic models were needed: three to compare the translation providers/methods (study 1) and one for the cross-country comparison of the climate change discourse (study 2). The first three models are based on the 633 differently translated Web pages. The fourth model is based on the entire corpus of 6,843 German, British and U.S. American Web pages.

For modeling, the structural topic model framework was used (STM; Roberts et al., 2016; Roberts et al., 2013).⁵ It builds on the same principles as the latent Dirichlet allocation (LDA; Blei et al., 2003)—the most common topic model in the social sciences—but allows users to incorporate additional information and covariates into the model (Roberts, Stewart, & Airoidi, 2016). These covariates can affect either the topical prevalence or the topical content. Topical prevalence refers to how much of a document is associated with a topic (e.g., German actors are more likely to speak about topic 1 than British actors). Topical content relates to the words used to discuss a topic (e.g., German-speaking actors are more likely than English speaking actors to use a particular word when they discuss topic 1). The advantage of the STM is that topical prevalence or topical content is not assumed to be constant across all

⁵ See Roberts et al. (2016), Roberts et al. (2013), and the R package’s vignette (Roberts, Stewart & Tingley, 2017) for a comprehensive description of the framework’s technical details.

documents, but rather may vary depending on its properties (Roberts et al., 2014). The STM framework, therefore, allows both the identification of topics and the estimation of their relationships to document properties.

Moreover, topical content covariates allow the model to “condition away systematic differences within the corpus that are not of primary interest” (Lucas et al., 2015, p. 263). It could be that the translation led to minor but systematic differences in the vocabulary of the documents. Since the primary interest is in the actors’ use of different topics based on their country of action (i.e., the big picture), small differences in the vocabulary should not influence the model. In this case, the inclusion of the documents’ original language as a topical covariate allowed the STM to capture systematic differences in the frequency a word was used by German and English speaking actors without affecting the overall outcome of the model. Technically this is done by defining “the distribution over the terms associated with the different topics as an exponential family model, similar to a multinomial logistic regression” (Roberts et al., 2016, p. 989). Thus, a topical content covariate indicating the document’s original language was included in the multinational model for the comparison of the climate change discourse, but not for the three models used for the comparison of the translation providers/methods.

The modeling procedure itself was done in *R* using the *stm* package (Roberts, Stewart & Tingley, 2017) and included the following steps. First, the optimal number of topics was evaluated. This was not a straightforward routine, as there is not a “right” number of topics for a given corpus or research question (Grimmer & Stewart, 2013, p. 285). For both studies several models were estimated, each with a different number of topics. To find the optimal number of topics, the models were then compared by three diagnostic indicators: their held-out likelihood (Wallach, Murray, Salakhutdinov, & Mimno, 2009), their residuals dispersion (Taddy, 2012), and their lower bounds. For the analysis of the climate change discourse,

however, diagnostic indicators were not the only criteria since the overall objective was to obtain the best solution in terms of interpretability (Maier et al., 2018). The models' outputs were thus also compared qualitatively by looking at the topics' top words (determined by their highest marginal probability). Considering diagnostic indicators and the models' overall interpretability, the 30-topics version of the model was ultimately chosen for study 2. For study 1, the 40-topics solution was chosen based on models' diagnostic indicators.

One challenge of topic models is that topics are abstract objects consisting of clusters of words that likely co-occur in the corpus's documents. To make use of them, they must be interpreted in theoretical terms (Jacobi, van Atteveldt, & Welbers, 2016). In the climate change case studied here, topics were examined through the theoretical lens of "same issue at the same time under similar criteria of relevance" (Eder & Kantner, 2000, p. 315, translation by author). The issue in this case is climate change. Thus, this study approaches topics as "criteria of relevance" of the climate change issue. This means that a topic theoretically represents a particular aspect of the climate change discourse that can be interpreted and named. It does not mean that word clusters necessarily represent a coherent position or frame, but they do represent at least some kind of sub-issue (Maier, Waldherr, Miltner, Jähnichen, & Pfetsch, 2017). However, not all topics are interpretable. Some do not represent a coherent concept or meaningful aspect of the debate and are thus hard to describe. Others are so-called boilerplate topics with no substantive meaning at all (Mimno & Blei, 2011). Such topics were excluded from the analysis because they do not help to answer the research questions. For the exclusion as well as for the labeling, the topic's top words and 10 randomly sampled documents with a relatively high probability (>0.6) of the topic were considered.

Label	Top words
Causes / effects of climate change	co2, atmosphere, carbon, increase, effect, ocean, temperature, greenhouse, gas, water
Climate change evidence	warm, global, temperature, chart, climate, co2, emission, year, trend, change
Climate change research	climate, research, change, science, university, institute, work, policy, study, impact
Climate modeling	climate, model, ipcc, prediction, study, predict, scientist, scientific, science, evidence
Climate politics	climate, change, government, green, emission, country, carbon, guardian, policy, action
Climate politics and science	science, climate, political, change, scientific, argument, make, claim, public, debate
Climate scepticism	solar, climate, radiation, watt, sun, wuwt, model, vapor, anthony, pinterest
Doubting climate research	paper, science, gore, publish, skeptic, write, review, medium, mann, michael
Economy and climate politics	company, industry, group, fund, oil, report, project, airport, business, plan
Energy consumption	energy, green, save, make, recycle, home, heat, reduce, waste, efficient
Energy sector	energy, gas, power, wind, fuel, coal, emission, price, cost, electricity
Environmental activism	wwf, work, change, campaign, climate, centre, live, business, people, sustainable
Extreme weather	climate, change, report, weather, research, publish, storm, impact, extreme, read
Food / health	food, science, movement, crop, environmental, ddt, permaculture, malaria, year, issue
Humanity	world, people, human, life, society, change, social, future, live, idea
Melting ice	ice, sea, arctic, level, rise, melt, ocean, year, glacier, polar
Science	physic, space, earth, science, scientist, planet, sun, nasa, system, year
Scientific results / concensus	warm, global, climate, change, scientist, report, temperature, year, world, ipcc
Temperature	temperature, warm, period, year, change, climate, record, global, trend, past
U.S. environmental politics	state, u.s, obama, american, epa, president, tax, federal, rule, regulation
U.S. fiscal policy	obama, spend, house, climate, democrat, year, president, republican, make, time
Wildlife protection	read, specie, climate, environment, fish, animal, green, wildlife, forest, energy

Table 1. Labels and top words (marginal highest probability) of the multinational climate change model.

The decision process was guided by two questions (Maier et al., 2018): (1) Does the topic depict a coherent (possibly controversial) aspect of the climate change discourse? (2) How can this aspect be described most comprehensively? If the top words and the documents pointed in a different direction or only one could be interpreted meaningfully, the topic was excluded. For the model used in study 3, eight topics were excluded, and 22 were labeled. Table 1 summarizes the labels and top words for the analysis of the online climate change discourse in Germany, the United Kingdom, and the United States. For this model, the German documents were translated as full-texts with DeepL.

Study 1: Robustness

Following de Vries et al. (2018), the vocabulary of the translated and preprocessed DTMs were compared to obtain a detailed impression of how much the results of the two

machine-translation providers differ and whether the translation of a DTM is a valid alternative to the full-text translation of a corpus. The first comparison focused on the level of documents. Using the cosine similarity, it was measured how similar the term vectors of a document are when the document was translated as full-texts with DeepL (DL/FT), as full-texts with Google Translate (GT/FT), and term-by-term with Google Translate (GT/DTM). Second, at the corpus level, the comparison of the unique terms gave an impression of both how comprehensive as well as how exclusive the vocabulary in the corresponding DTMs is.

De Vries et al. (2018) further assessed the similarity of translated corpora by comparing the results of topic models computed on the translated corpora. This makes sense, since it ultimately depends on the similarity of the topic models whether the conclusions drawn from the corpora are the same. Therefore the focus here was primarily on the topics' interpretations as well as on their prevalence in the documents.

DL/FT	GT/FT	GT/DTM
ipcc	ipcc	climate
change	co2	ipcc
catastrophe	catastrophe	lie
co2	lie	co2
lie	state	world
state	world	catastrophe
world	people	front
people	report	year
year	earth	kyoto
panel	year	people

Table 2. Top words (marginal highest probability) of matched topics describing climate change as lie.

The first step was to see which topics can be matched with each other. To do so, the probability values of 30 top words were compared all labeled topics. Building on Niekler and Jähnichen's (2012) approach to match topic model results, the cosine similarity was used to measure the topics' similarity. All topic pairs with a relatively high cosine similarity (>0.4) were then examined in depth. This means that the previously assigned labels as well as the

top words of the topics were compared qualitatively with each other. If two topics described the same thematic aspect of the climate change discourse, they were considered a match.

Table 2 shows an example for three matched topics. When looking at the top words, it can be stated that they describe the same thematic aspect of the discourse. This impression is confirmed by the measured cosine similarities. The resulting similarity of 0.859 for the Google Translate (GT/FT) and the DeepL (DL/FT) translations indicates a very good match. The similarity of the GT/DTM's topic with the others, however, not quite as good. The cosine similarity is 0.565 for the match with the DL/FT's topic and 0.626 for topic of the GT/FT model.

The second step focused on the bigger picture by looking at the number of matched topics as well as at the average cosine similarity. The more topics could be matched, the more similar two models are. Ideally, each topic could only be matched with one topic from the other model (unique match). In practice, however, a topic was often matched multiple times (multi-match). The interpretation of such multi-matches is difficult, as they indicate both similarity and difference of the compared models at the same time. Since the reason for multi-matches lies not only in different corpora (due to different translations), but also in the topic models' generative process (cf. Maier et al., 2018), only those pairs out of the multi-matches with the highest cosine similarity (first matches) as well as the unique matches were considered in the final step of the study.

This step concentrated on the question whether the matched topics have the same prevalence in the documents of the translated corpora. Or in other words: Is the distribution of the matched topics the same across the documents? This question is central, since most studies using topic models just rely on the topic proportions to draw conclusions from a corpus. As suggested by de Vries et al. (2018), the matched topics' distributions over all documents were correlated with each other to determine their similarity.

Study 2: Integrability

To measure the discourse convergence of the climate change discourses in Germany, the United Kingdom, and the United States, the multinational topic model was analyzed in two ways: (1) a comparison of the mean topic proportions per country, (2) a comparison of the relative frequency of topics as the top topics per country. The top topic is the topic with the highest average probability over all Web pages of an actor. In the first case, differences between countries were tested for significance using ANOVA. In the second case, Fisher's exact test was used due to the occasionally small number of cases in which a topic was the top topic.

The topic model results are generally discussed directly at the individual document level (i.e., Web page level) because they are the main unit for the calculation of the model (i.e., "bag of words"). In this study, however, the question is whether actors in Germany, the United Kingdom, and the United States discuss the same aspects of climate change on their websites. Thus, the focus is on the actors' websites rather than on the individual Web pages. In this case, the actor definition includes all individuals or organizations who are represented in the sample with at least one Web page. Actors can therefore be political actors (e. g. governments, offices, parties), civil society actors (e.g., environmental organizations, research institutions, churches, foundations, citizens' initiatives, blogs), media, or companies. For the analysis, the individual Web pages were aggregated to actors based on the manual coding of the "about us" pages. The aggregation was done by calculating the mean topic probability over all Web pages of an actor. In contrast to summing up the topic probabilities, calculating the mean maintains the bag-of-words logic on the actor level and therefore allows a direct comparison of actors further in the analytical process.

Results

Study 1: Robustness

The first comparison to assess whether the different translation providers/methods produce similar results is that of the translated documents. For each document, the cosine similarity between the word vectors (i.e., “bag of words”) of the different translated versions was calculated. For two identical translations, the cosine similarity would be 1. If no words in both translations are identical, the similarity is 0.

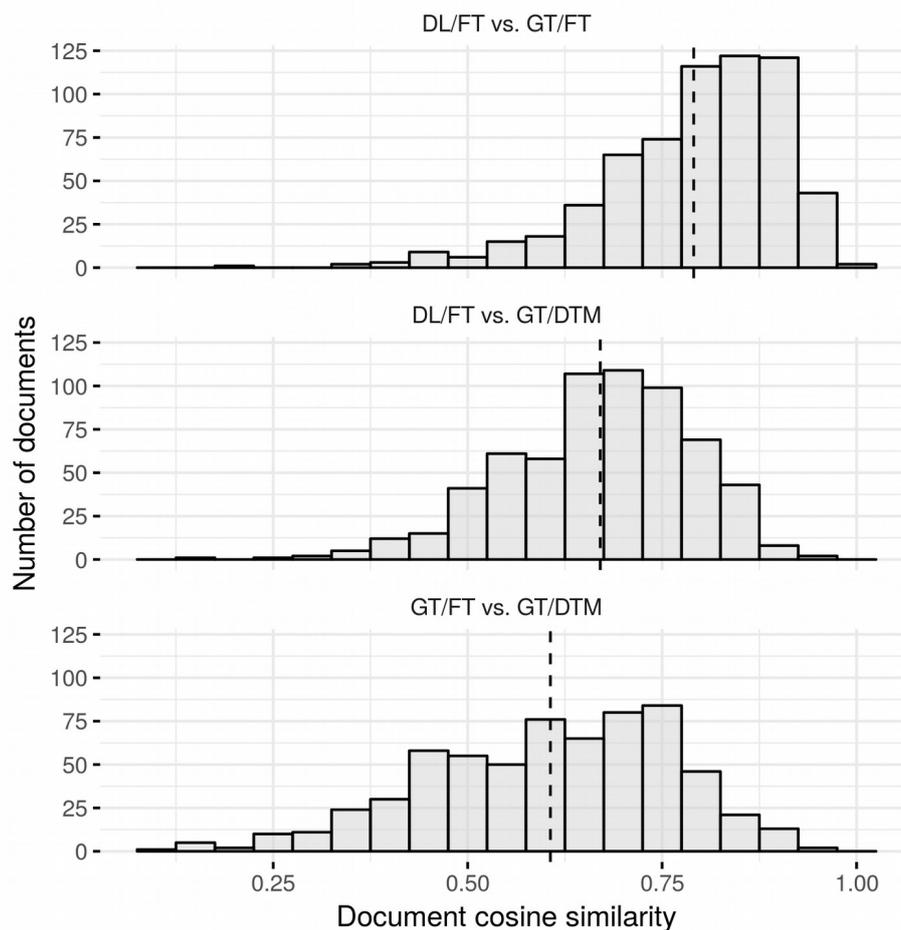


Figure 1. Distributions of cosine similarity between documents per translation method pair.

Figure 1 shows the distribution of cosine similarities between document pairs ($n = 633$) for every translation method pair. The figure shows that Google Translate (GT/FT) and DeepL (DL/FT) produce quite similar results when whole Web pages are translated ($M = 0.791$, $SD = 0.117$). However, if only the single words of the DTM were translated

(GT/DTM), the similarities between the documents are smaller, although still respectable. An ANOVA showed that the differences between the pairings' distributions are significant ($F(1, 1897) = 216.6, p < 0.001$).⁶ Interestingly, the results of the DL/FT and the GT/DTM translation are closer together ($M = 0.670, SD = 0.120$) than the GT/FT and the GT/DTM translation ($M = 0.606, SD = 0.158$), even though the DTM translation was also done with Google Translate.

A second way to compare the translations on a corpus level is to compare the vocabulary in the respective DTMs. A larger vocabulary means that thematic structures are represented in a more detailed and thus probably more accurate way. Furthermore, if large parts of the vocabulary match, the texts have been translated correspondingly. As shown in Figure 2, the vocabularies in the full-text translations are slightly larger than in the term-by-term translation of the DTM. The full-text translation with DeepL (DL/FT) produced the vocabulary with the most unique terms (features; 7,372), followed by the full-text translation with Google Translate (GT/FT; 6,978). The term-by-term translation with Google Translate (GT/DTM) produced the result with the smallest number of unique terms (6,843). The number of overlapping features is also larger for the two full-text translations (6,161 terms) than for the full-text translations versus term-by-term translation (GT/FT vs. GT/DTM: 5,719; DL/FT vs. GT/DTM: 5,533 terms). Not surprisingly, but other than above, the results of the GT/FT and the GT/DTM translations are closer to each other than to the DL/FT translation. However, the DeepL translation is the most detailed and thus probably the most accurate.

6 The documents at the lower end of the left tails are mainly longer blog posts with many special characters and unconventional punctuation. In such cases, sentence decomposition often led to sentence fragments, which were then translated differently by the translation services.

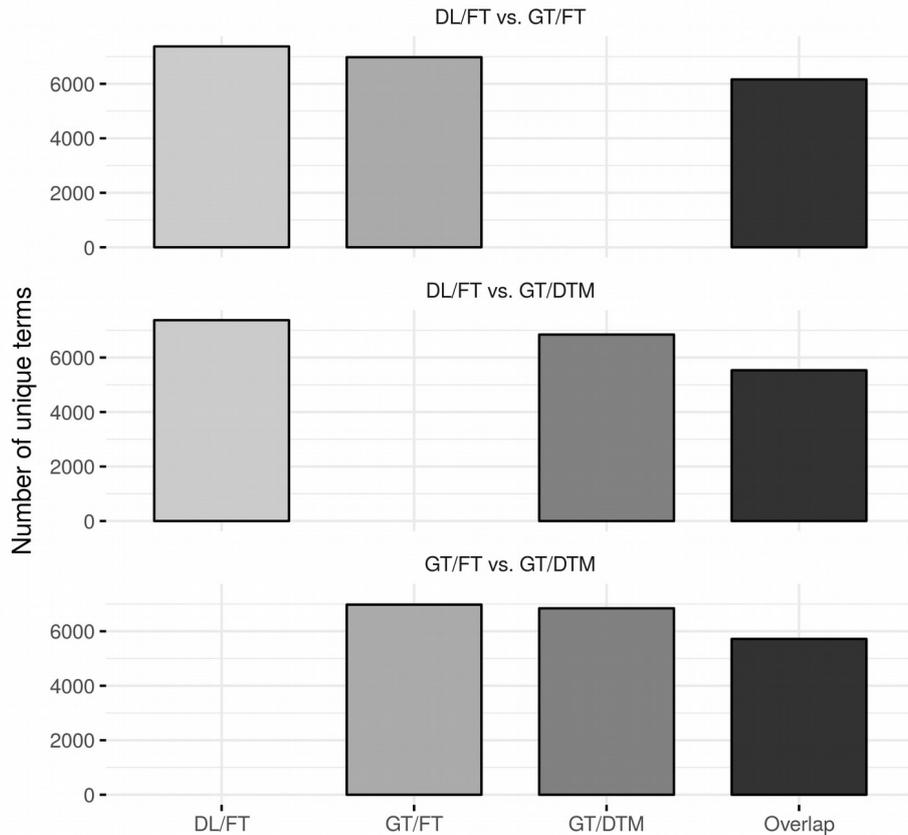


Figure 2. Unique terms in the translated and preprocessed DTMs and their overlap.

The apparent reason for the smaller vocabulary of the GT/DTM translation is that words with several meanings (depending on the context) are lumped together and then translated with one and the same term. Part of the semantic information is therefore lost in the process of a DTM translation. Whether this causes major divergences in the corpus can be checked by comparing the results of the topic models.

The first comparison of the topic model results denotes to the number of matched topics. For every translated corpus a model with 40 topics was computed. The models were then interpreted as described above. For the DL/FT model a total of 29 topics could be interpreted and labeled based on their top words and a selection of documents. The same was true for a total of 28 topics in the GT/FT model and 29 topics in the GT/DTM model. Thus, there were no notable differences in the interpretability of the three models.

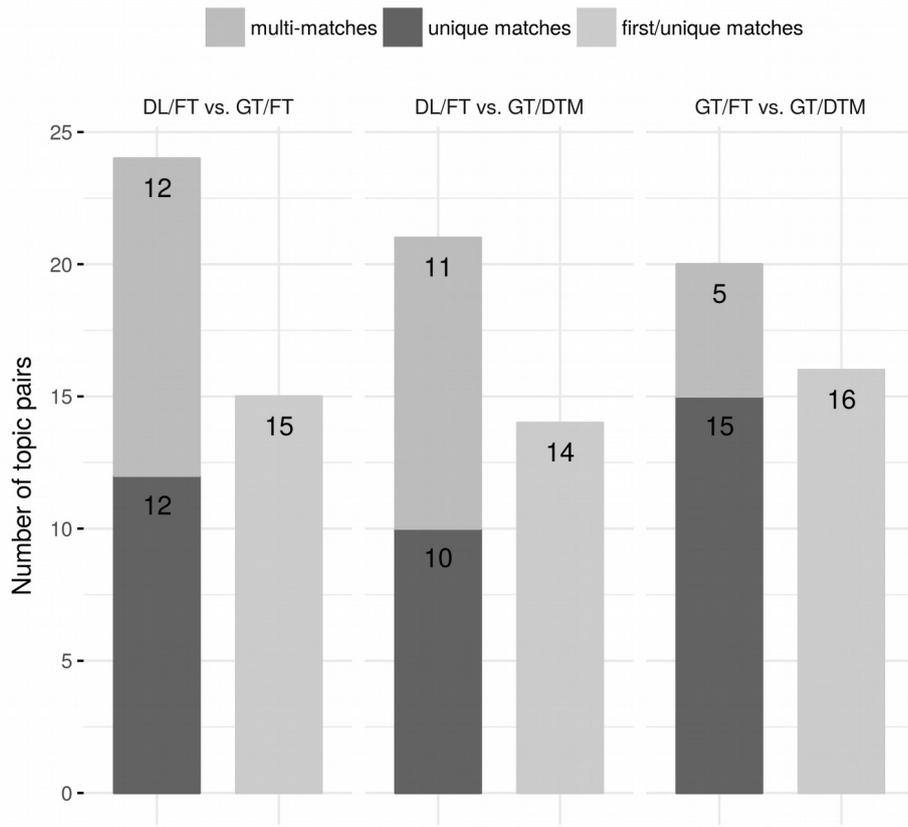


Figure 3. Matched topics.

Of the 29 respectively 28 topics a total of 24 could be matched in the DL/FT versus GT/FT comparison. However, 12 pairings were multi-matches (i.e., one or both topics of the pair is also involved in another match). If only the pairings with the best cosine similarity (first matches) as well as the unique matches were considered, 15 topics could be matched. For the DL/FT versus GT/DTM comparison a total of 21 topics could be matched, but only 14 as first/unique matches. Finally, for the GT/FT versus GT/DTM comparison 20 topics could be matched, 16 of them as first/unique matches. Figure 3 illustrates these findings.

If the multi-matches are included in the picture, a considerable part of the interpretable topics could be matched in all three cases. Comparing the two full-text models (DL/FT vs. GT/FT), only 4 respectively 5 topics have no equivalent in the other model. No matter which of the two models is used, the overall picture regarding topical content is very similar. Not quite as good is the result when comparing the full-text models with the GT/DTM model. However, in both cases more than two thirds of the labeled topics could be

matched. The majority of the topics therefore point in the same substantive direction. If only the first matches are considered, the differences between the models vanish, as only about half of the topics could be matched in all three comparisons.

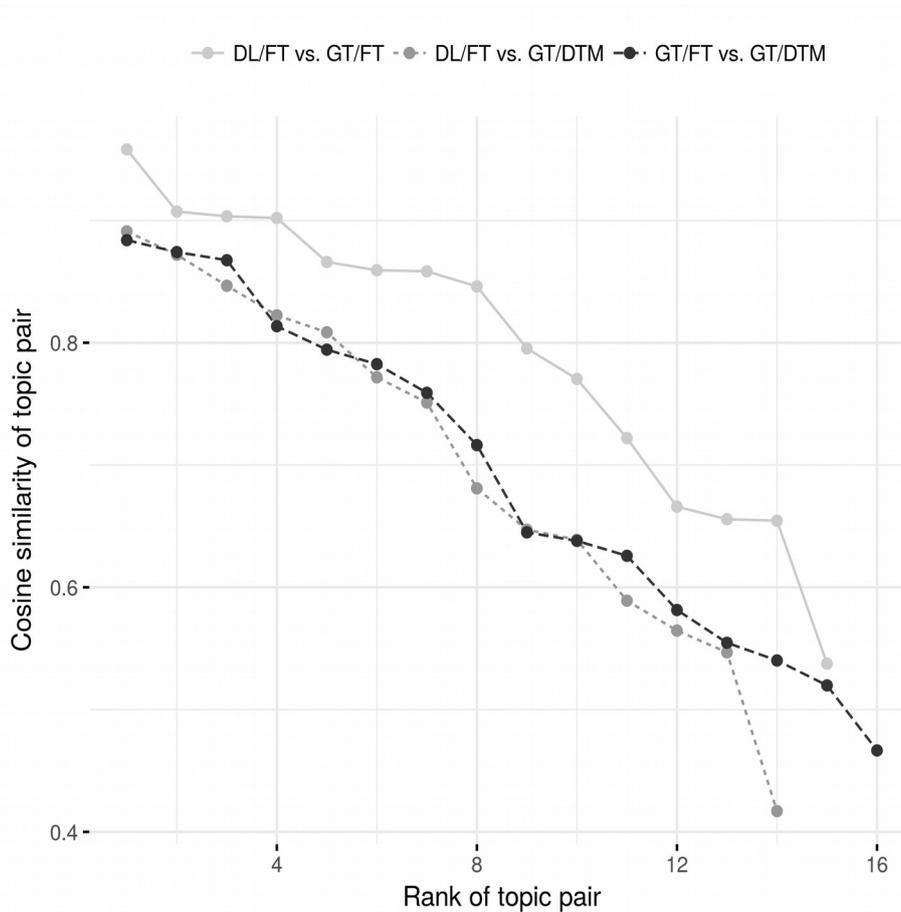


Figure 4. Cosine similarity of matched topics (first matches).

When looking at the first matches more closely, however, it becomes clear that those of the full-text models (DL/FT vs. GT/FT) have a higher mean cosine similarity ($M = 0.793$, $SD = 0.121$) than when they are compared to the GT/DTM model (DL/FT vs. GT/DTM: $M = 0.703$, $SD = 0.142$; GT/FT vs. GT/DTM: $M = 0.691$, $SD = 0.137$). Figure 4 illustrates this. With regard to topical content it can therefore be said: The two full-text models are more similar to each other than they are to the GT/DTM model.

But are the matched topics equally important for the same documents? To measure this, the individual topics' proportions over all Web pages were correlated per translation method pair. Figure 5 shows the overall picture. The descriptive figures are summarized in

Table 3. With mean correlation coefficients between 0.6 and 0.7 in all three cases, most matched topics correlate quite well. Although the resulting coefficients are wider distributed in the comparisons of the full-text models with the GT/DTM model, the results do not differ significantly from the comparison of the two full-text model (ANOVA: $F(1, 43) = 0.405$, $p = 0.569$). Looking at the topical prevalence of the topics, it can thus be said that the conclusions drawn from the different models would be quite similar.

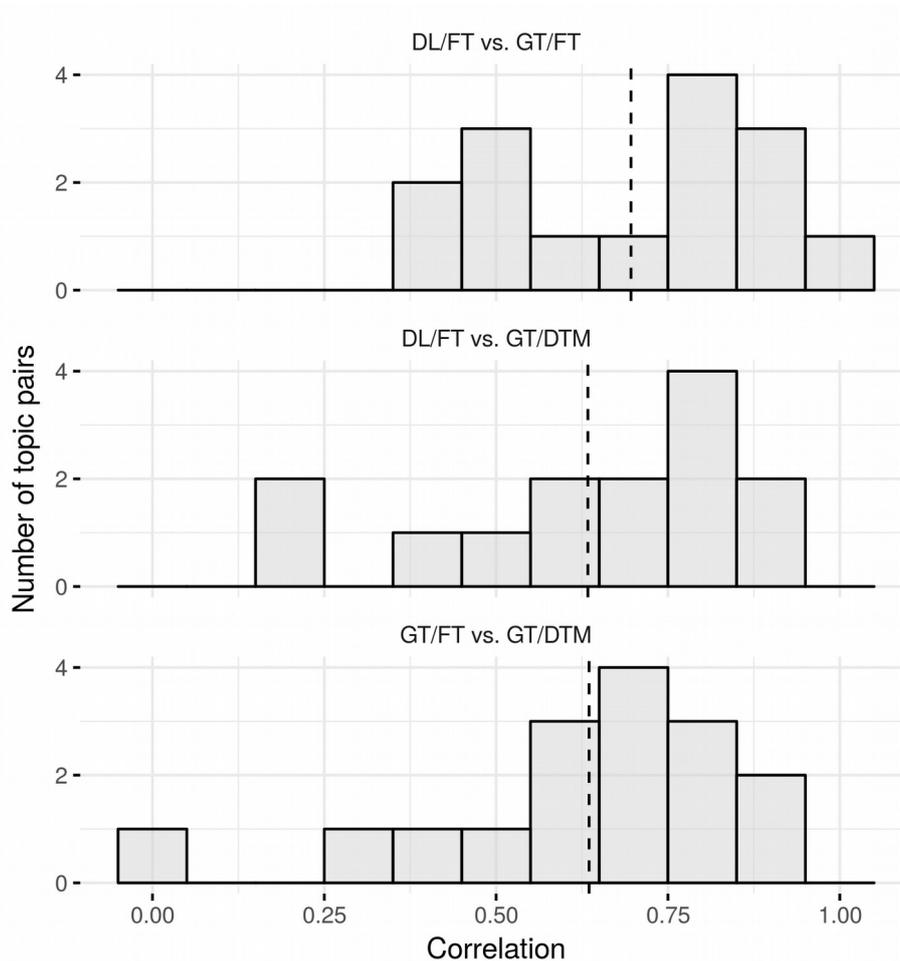


Figure 5. Topical prevalence correlations.

Translation method pair	N	Mean	SD	Min	Max
DL/FT vs. GL/FT	15	0.696	0.209	0.384	0.990
DL/FT vs. GL/DTM	14	0.633	0.231	0.174	0.907
GL/FT vs. GL/DTM	16	0.635	0.233	0.032	0.922

Table 4. Descriptive figures for topical prevalence correlations.

To sum up, the differences between the two full-text translations are small. The accuracy of DeepL and Google Translate seems to be similar for full-text translations. Due to the larger vocabulary, DeepL may be slightly more precise than Google Translate, but it is safe to say that the choice of translation service plays a minor role. More important is the choice of the translation method, as the differences between the full-text translations and the DTM translation are bigger than between the full-text translations. However, the differences are not of a fundamental nature. Both the majority of documents as well as topics point in the same substantive direction for DTM and full-text translations. The conclusion of Lucas et al. (2015) can therefore be confirmed that two researchers using a DTM and a full-text translation would reach the same substantive conclusions. Thus, the translation of the individual terms of a DTM can be a useful shortcut for the translation of larger corpora. Wherever possible, however, the whole texts should be translated. The smaller vocabulary of the DTM translation is a clear indicator that information is lost due to the previous preprocessing of the documents. The size of the vocabulary is also the reason, why the DeepL translation was used for the following analysis of the climate change discourse.

Study 2: Integrability

The first comparison to measure whether actors in Germany, the United Kingdom, and the United States discuss climate change “under similar criteria of relevance” uses the mean topic proportions. Figure 6 gives an overview of the topic proportion based on the mean topic proportions of the actors. Looking at the German actors, it is noticeable that climate change research plays an important role in their communication. Not only because the corresponding topic has by far the highest probability, but also because other “scientific” topics, such as the causes and effects of climate change, the melting of ice, the temperature, or scientific results have high mean values. Other topics frequently encountered by German actors include environmental activism as well as economic issues (especially the energy

industry). U.S. American actors also give particular weight to scientific aspects of the climate change issues (causes and effects of climate change, climate change research, the temperature as well as scientific results, and the scientific consensus). Other important sub-issues are the U.S. environmental politics and the energy sector. British actors, finally, tend to place more emphasis on politics. Topics such as climate politics, economy and climate politics, the (regulation of) the energy sector as well as environmental activism have remarkably high average proportions. Climate research also plays a crucial role for British actors, albeit not a very important one.

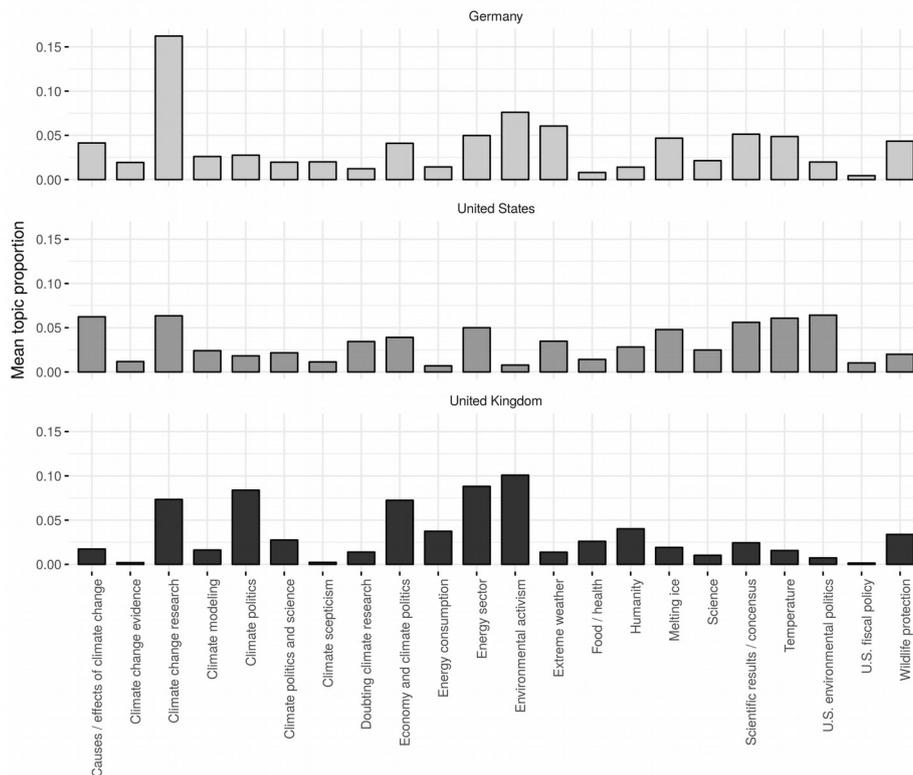


Figure 6. Mean topic proportion (over all actors).

With regard to differences between the countries, an ANOVA shows that most of the expected topic probabilities are actually different (see Appendix B for a table with the test results). British actors emphasize political as well as economic aspects (climate politics, environmental activism, economy and climate politics, energy sector) significantly more than actors in Germany and the United States. Climate science, on the other hand, appears to be a

rather German sub-issues (climate change research, extreme weather). Not surprisingly, U.S. environmental politics is on average more important to U.S. actors than to others.

A clearer picture results if only the top topics are considered that have the highest mean proportion over all Web pages of an actor. This simplification is helpful, as each document theoretically consists of each topic. Figure 7 shows the relative frequency of topics as the top topics for the three countries. Germany again shows the highest interest in climate change research. The impression gained before is also confirmed by the British actors, whose main concern is often political in nature (climate politics, energy sector, environmental activism, economy and climate politics, humanity). Particularly interesting is the picture that arises for U.S. actors. Clearer than in the analysis of the mean values, the top topics show a broad interest in fundamental questions on climate change. This holds true for topics such as causes and effects of climate change, scientific results and consensus as well as the temperature (trend). A substantial proportion of the actors are mainly concerned with U.S. environmental politics.

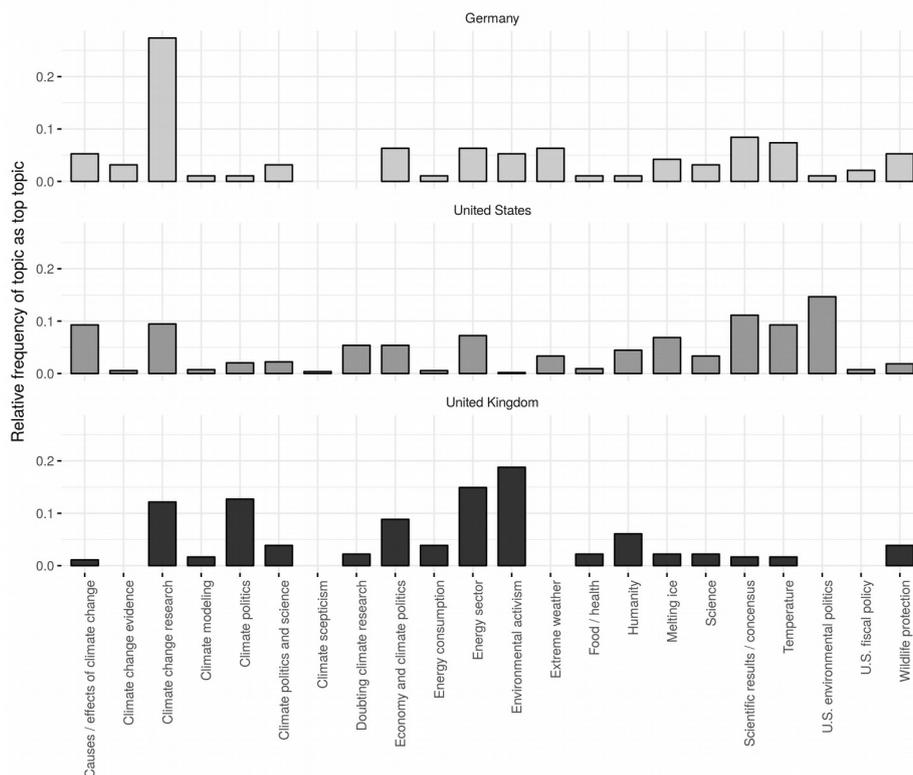


Figure 7. Relative frequency of topics as top topic.

Not every topic that is typical of a country's discourse is also a unique feature. The sub-issue of economy and climate politics is relatively important in the United Kingdom, but it is not more important than in the two other countries. However, using Fisher's exact test (see Appendix C for a table with all the test results), some differences in the distribution can be identified. According to the test results, climate change research can be defined as a German topic, whereas climate politics, the energy sector, and environmental activism appear to be typically British. Typical U.S. American topics are the causes and effects of climate change, the scientific consensus, and the U.S. environmental politics. Nevertheless, the public discourses about climate change in Germany, the United Kingdom, and the United States are not fundamentally different. Since there are very few topics (U.S. environmental politics, climate skepticism) that solely belong to actors of only one country, the observed differences relate mostly to the salience with which a topic is discussed, and less to the question of whether a topic is discussed at all. Actors in both countries therefore discuss climate change "under similar criteria of relevance" (Eder & Kantner, 2000, p. 315), but they do not always attribute the same importance to them. It can therefore be said that the online climate change discourse in Germany, the United Kingdom, and the United States partly converges, but each of them also has its clear national characteristics.

Discussion

The availability of affordable and quickly improving machine-translation services introduces new opportunities for comparative communication research. With online translation services such as Google Translate and DeepL, multilingual text corpora can be easily transformed into monolingual corpora, which can then be analyzed with, for example, topic models. In this paper, the potential of such a combined approach has been assessed, both from a methodological and substantive point of view.

From a methodological point of view, it can be said that when texts are translated for an analysis with topic models, similar results can be obtained no matter which translation service (Google Translate vs. DeepL) or method (full-text vs. DTM) is used. Looking at the big picture—and given that the topic modeling procedure is kept unchanged—the combination of topic models and machine translation therefore appears to be quite robust. Looking closer, however, it turns out that the choice of translation service is less of a factor than the choice of method. While the full-text translations of Google Translate and DeepL are relatively similar, they differ more from the translation of the individual terms of a DTM. This shows that machine translation services, just like humans, use the context in which a word is written in a text to determine its exact meaning. Nevertheless, the comparisons have shown that the results of the DTM translation are not too far from those of the full-text translations. To translate the individual terms of a DTM is thus an acceptable and cost-effective alternative to obtain a monolingual corpus from a multilingual one. For better results, however, the first choice should always be the translation of entire documents.

To examine whether topic modeling coupled with machine translation can be linked to existing theories and work from the field of comparative communication research (i.e., the substantive point of view), the online climate change discourses in Germany, the United Kingdom and the United States of America was compared with regard to their transnationality. For the analysis, the German Web pages were translated with DeepL as a whole. The results indicate that there are parallels as well as divergences between the debates in the countries. This is consistent with what is known about the climate change discourse in offline mass media. Ivanova (2017) has shown that the media agenda for climate change is similar in all three countries, but there are differences in the importance of different topics. Similarities to the findings described here are particularly apparent in the greater emphasis on scientific descriptions in Germany and the stronger accentuation of political topics in the

United Kingdom (cf. Ivanova, 2017). One potential explanation for the different national priorities as well as for potential parallels between national online and offline agendas could be the national political agenda (cf. Grundman & Scott, 2014). But it must be left to others to investigate this further.

This study showed that machine translation and topic models are a powerful couple when it comes to the analysis of multilingual corpora. Even in large amounts of text, a combined approach makes it possible to identify and compare thematic structures across language boundaries with relative ease. The cross-sectional comparison of two cases made here, is of course only one possible application. Longitudinal comparisons, as well as comparisons between several cases (i.e., languages, countries, and document types), are also possible. Moreover, content information obtained from topic models can also be combined with other information, such as network data. This could contribute to a better understanding of underlying discursive mechanisms, especially in the case of online communication. Topic models coupled with machine translation thus is a valuable addition to the toolbox of comparative communication researchers.

However, a combined use topic models and machine translation is not a cure-all. First, topic models—with or without translation—are hardly the right tool for in-depth analyses of discourses. Their strength is the detection of relatively coarse thematic structures in large text corpora, but they are blind to more complex thematic structures, such as sentiments or arguments. Thus, if the focus is on more complex linguistic structures, other, possibly manual methods of content analysis are required. Second, topic models must be put on a solid theoretical foundation in order to be interpreted meaningfully.

Third, it must always be asked whether the analyzed texts should be sent to servers of online translation services. When public communication is studied, as here, this is usually not a problem, but when it comes to private communication, such as interviews, letters or e-

mails, protecting the privacy of the people involved must be an issue. In such cases, online translation services should be avoided.

With regard to the two studies presented here, there are further limitations which must be addressed. In particular, it is not possible to determine whether the different translations providers/methods have caused systematic biases (e.g. by translating into British/U.S. American English). The validity of the translation can not be assessed with the corpus of Web documents used here, as there is no benchmark available. The same applies for the accuracy of the translation. It is not certain that the larger vocabulary of the DL/FT translation actually means a more accurate translation, although a direct but unsystematic comparison of randomly selected documents from both full-text translations supports this assumption. In order to be sure, however, a comparison with a reference translation (“gold standard”) is necessary (cf. de Vries et al., 2018).

Another restriction relates to the languages studied here. German to English translations are believed to be quite good in comparison to translations between other languages, as they are both Germanic languages and because there is much training data available (e.g. human-translated transcripts of parliament debates). Although Lucas et al. (2015) have reported similar findings for documents written in Chinese and Arabic, further studies have to show whether the results reported here are also valid for other languages.

It should also be borne in mind that other languages may pose other challenges—such as the ngrams in the German to English translation done here—and that the proposed procedure may therefore require some adjustments if it is used with other languages. It is also important to note again that changes in the preprocessing as well as in the modeling procedure may well affect the results (cf. Maier et al., 2018).

The final restriction relates to the types of models that can be reliably computed with machine-translated texts. Possible inaccuracies caused by the machine translation are less of a

concern with bag-of-words topic models due to the relatively rigid preprocessing procedure and their focus on coarse semantic structures. Whether machine translation is also useful for models that focus on finer semantic structures (e.g., arguments), and thus depend on correct syntactical structures, is uncertain. Therefore, future research should also focus on the usefulness of machine translation for linguistically more sophisticated methods.

References

- Adam, S., Häussler, T., Schmid-Petri, H., & Reber, U. (2016). Identifying and Analyzing Hyperlink Issue Networks. In G. Vowe & P. Henn (Eds.), *Political Communication in the Online World: Theoretical Approaches and Research Designs* (pp. 233–247). New York, NY: Routledge.
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O. & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Language in Social Media* (p. 30–38). Retrieved from <http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf>
- Balahur, A. & Turchi, M. (2014). Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Computer Speech and Language*, 28, 56–75. doi:10.1016/j.csl.2013.03.004
- Bennett, W. L., Lang, S. & Segerberg, A. (2015). European issue publics online: The cases of climate change and fair trade. In T. Risse (Ed.), *European public spheres: Politics is back*. Cambridge: Cambridge University Press.
- Benoit, K., Schwarz, D. & Traber, D. (2012, June). *The sincerity of political speech in parliamentary systems: A comparison of ideal points scaling using legislative speech and votes*. Presented at the Second Annual Conference of the European Political Science Association (EPSA), Berlin. Retrieved from http://www.dscharz.ch/downloads/EPSA_2012_BenoitSchwarzTraber.pdf?attredirects=0
- Benoit, K., Watanabe, K., Nulty, P., Obeng, A., Wang, H., Lauderdale, B. & Lowe, W. (2017). *quanteda: Quantitative analysis of textual data*. R package version 0.99. <http://quanteda.io>
- Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4/5), 993–1022.

- Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, 4(1), 8–23. doi:10.1080/21670811.2015.1096598
- Carter, N. (2014). The politics of climate change in the UK. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3), 423–433. doi:10.1002/wcc.274
- de Vries, E., Schoonevelde, M., & 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. doi:10.1017/pan.2018.26
- DeepL. (2017). *Der DeepL Übersetzer im Vergleich zur Konkurrenz*. Retrieved from <https://www.deepl.com/press.html>
- Denny, M. J. & Spirling, A. (2017). *Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it*. Retrieved from <https://ssrn.com/abstract=2849145>
- Eder, K. (2000). Zur Transformation nationalstaatlicher Öffentlichkeit in Europa: Von der Sprachgemeinschaft zur issuespezifischen Kommunikationsgemeinschaft. *Berliner Journal für Soziologie*, 10, 167-284. doi:10.1007/BF03204349
- Eder, K. & Kantner, C. (2000). Transnationale Resonanzstrukturen in Europa: Eine Kritik der Rede vom Öffentlichkeitsdefizit. In M. Bach (Ed.), *Die Europäisierung nationaler Gesellschaften* (p. 306–331). Wiesbaden: Westdeutscher Verlag.
- Elgesem, D., Steskal, L., & Diakopoulos, N. (2015). Structure and content of the discourse on climate change in the blogosphere: The big picture. *Environmental Communication*, 9(2), 169–188. doi:10.1080/17524032.2014.983536
- Esser, F. (2013). The emerging paradigm of comparative communication enquiry: Advancing cross-national research in times of globalization. *International Journal of Communication*, 7, 113–128.

- Esser, F. & Hanitzsch, T. (2012). On the why and how of comparative inquiry in communication studies. In F. Esser & Hanitzsch (Eds.), *Handbook of comparative communication research* (p. 3–22). New York, NY: Routledge.
- Grimmer, J. & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
doi:10.1093/pan/mps028
- Grundmann, R. & Scott, M. (2014). Disputed climate science in the media: Do countries matter? *Public Understanding of Science*, 23(2). doi:10.1177/0963662512467732
- Hampshire, S. & Porta Salvia, C. (2010). Translation and the Internet: Evaluating the quality of free online machine translators. *Quaderns: Revista de Traducció*, 17, 197-209.
Retrieved from <https://www.raco.cat/index.php/QuadernsTraduccio/article/view/194256>
- Häussler, T., Adam, S., Schmid-Petri, H. & Reber, U. (2017). How political conflict shapes online spaces. A comparison of climate change hyperlink networks in the U.S. and Germany. *International Journal of Communication*, 11, 3096–3117. Retrieved from <http://ijoc.org/index.php/ijoc/article/view/5644>
- Ivanova, A. (2017). *Transnationalisierung von Öffentlichkeit. Eine länderübergreifende Langzeitanalyse der Klimaberichterstattung in Leitmedien*. Wiesbaden, DE: Springer Fachmedien.
- Ivanova, A., Schmidt, A., & Schäfer, M. S. (2014). Global climate change, global public sphere? Media attention for climate change in 27 countries. In M. José Canel & K. Voltmer (Eds.), *Comparing political communication across time and space: New studies in an emerging field*. New York, NY: Palgrave Macmillan.
- Jacobi, C., van Atteveldt, W. & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89-106.
doi:10.1080/21670811.2015.1093271

- Jang, S. M. & Hart, P. S. (2015). Polarized frames on “climate change” and “global warming” across countries and states: Evidence from Twitter big data. *Global Environmental Change*, 32, 11–17. doi:10.1016/j.gloenvcha.2015.02.010
- Kantner, C. (2004). *Kein modernes Babel: Kommunikative Voraussetzungen europäischer Öffentlichkeit*. Wiesbaden, DE: VS Verlag für Sozialwissenschaften.
- Kirilenko, A. P. & Stepchenkova, S. O. (2014). Public microblogging on climate change: One year of Twitter worldwide. *Global Environmental Change*, 26, 171–187. doi:10.1016/j.gloenvcha.2014.02.008
- Livingstone, S. (2003). On the challenges of cross-national comparative media research. *European Journal of Communication*, 18(4), 477–500. doi:10.1177/0267323103184003
- Livingstone, S. (2012). Challenges to comparative research in a globalizing media landscape. In F. Esser & T. Hanitzsch (Eds.), *Handbook of comparative communication research* (pp. 415-429). London, UK: Routledge.
- Lotz, S. & 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. doi:10.5842/43-0-205
- Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A. & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23(2), 254–277. doi:10.1093/pan/mpu019
- Lucas, C. & Tingley, D. (2014). *translateR: Bindings for the Google and Microsoft translation APIs*. R package version 1.0. <https://CRAN.R-project.org/package=translateR>
- Lück, J., Wessler, H., Wozniak, A., & Lycarião, D. (2016). Counterbalancing global media frames with nationally colored narratives: A comparative study of news narratives and news framing in the climate change coverage of five countries. *Journalism*. Advance online publication. doi:10.1177/1464884916680372

- Maier, D., Waldherr, A., Miltner, P., Jähnichen, P. & Pfetsch, B. (2017). Exploring issues in a networked public sphere: Combining hyperlink network analysis and topic modeling. *Social Science Computer Review, online first*, 1–18. doi:10.1177/0894439317690337
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H. & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2–3), 93-118. doi:10.1080/19312458.2018.1430754
- Marres, N. & Rogers, R. (2005). Recipe for tracing the fate of issues and their publics on the Web. In B. Latour & P. Weibel (Eds.), *Making things public: Atmospheres of democracy*. Karlsruhe, Cambridge: ZKM Center for Art and Media, The MIT Press.
- Mimno, D. & Blei, D. (2011). Bayesian checking for topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (p. 227–237), Edinburgh, UK. Retrieved from <http://www.cs.columbia.edu/~blei/papers/MimnoBlei2011.pdf>
- Mimno, D., Wallach, H., Naradowsky, J., Smith, D. & McCallum, A. (2009). Polylingual topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2* (p. 880–889), Singapore, SG. Retrieved from <http://dl.acm.org/citation.cfm?id=1699571.1699627>
- Niekler, A. & Jähnichen, P. (2012). Matching Results of Latent Dirichlet Allocation for Text. In *Proceedings of ICCM 2012, 11th International Conference on Cognitive Modeling* (p. 317-322). Retrieved from <http://asv.informatik.uni-leipzig.de/publication/file/210/nieklerjaehnichenICCM2012.pdf>

- O'Neill, S., Williams, H. T., Kurz, T., Wiersma, B., & Boykoff, M. (2015). Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Climate Change*, 5, 380–385. doi:10.1038/NCLIMATE2535
- Olausson, U. (2009). Global warming—Global responsibility? Media frames of collective action and scientific certainty. *Public Understanding of Science*, 18, 421–436. doi:10.1177/0963662507081242
- Painter, J., & Ashe, T. (2012). Cross-national comparison of the presence of climate scepticism in the print media in six countries, 2007-10. *Environmental Research Letters*, 7(4), 1–8. doi:10.1088/1748-9326/7/4/044005
- Pennings, P. (2011). Assessing the ‘gold standard’ of party policy placements: Is computerized replication possible? *Electoral Studies*, 30(3), 561–570. doi:10.1016/j.electstud.2011.05.001
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Roberts, M. E., Stewart, B. M. & Airoidi, E. M. (2016). A model of text for experimentation in social sciences. *Journal of the American Statistical Association*, 111(515). doi:10.1080/01621459.2016.1141684
- Roberts, M. E., Stewart, B. M. & Tingley, D. (2017). *stm: R package for Structural Topic Models*. <http://www.structuraltopicmodel.com>
- Roberts, M. E., Stewart, B. M., Tingley, D. & Airoidi, E. M. (2013). *The structural topic model and applied social science*. Presented at the Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation. Retrieved from <http://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf>

- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Kushner Gadarian, S., Albertson, B. & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082.
doi:10.1111/ajps.12103
- Rogers, R. (2013). Mapping public Web space with the Issuecrawler. In B. Reber & C. Brossaud (Eds.), *Digital cognitive technologies: Epistemology and the knowledge economy*. Hoboken, NJ: Wiley.
- Salton, G. (1991). Developments in automatic text retrieval. *Science*, 253(5023), 974–980.
- Schäfer, M. (2012). Online communication on climate change and climate politics: A literature review. *WIREs Climate Change*, 3(6), 527–543. doi:10.1002/wcc.191
- Schmid-Petri, H., & Arlt, D. (2016). Constructing an illusion of scientific uncertainty? Framing climate change in German and British print media. *Communications*, 41(3), 265–289. doi:10.1515/commun-2016-0011
- Schmidt, A., Ivanova, A. & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change*, 23(5), 1233–1248. doi:10.1016/j.gloenvcha.2013.07.020
- Sharman, A. (2014). Mapping the climate sceptical blogosphere. *Global Environmental Change*, 26, 159–170. doi:10.1016/j.gloenvcha.2014.03.003
- Taddy, M. A. (2012). On estimation and selection for topic models. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics* (p. 1184–1193). Retrieved from <http://proceedings.mlr.press/v22/taddy12/taddy12.pdf>
- Wallach, H., Murray, I., Salakhutdinov, R. & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning* (p. 1105–1112), Montreal, CA. Retrieved from <http://dirichlet.net/pdf/wallach09evaluation.pdf>

Wessler, H., Peters, B., Brüggemann, M., Kleinen-von Königslöw, K. & Sift, S. (2008).

Transnationalization of public spheres. Basingstoke, UK: Palgrave Macmillan.

Wessler, H., Wozniak, A., Hofer, L., & Lück, J. (2016). Global multimodal news frames on climate change: A comparison of five democracies around the world. *The International Journal of Press/Politics*, 21(4), 423–445. doi:10.1177/1940161216661848

Zhou, Y., Cristea, A. I. & Roberts, Z. (2015). Is Wikipedia really neutral? A sentiment perspective study of war-related Wikipedia articles since 1945. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation* (p. 160-168), Shanghai, CN. Retrieved from <http://www.aclweb.org/anthology/Y15-1019>

Appendix

Appendix A: Webcrawler Starting Points

Starting points for the snowball-sampling of websites in Germany, the United Kingdom, and the United States.

Germany		
Advocates	Heinrich Boell Stiftung	http://klima-der-gerechtigkeit.boellblog.org
	Greenpeace Germany	http://www.greenpeace.de/themen/klima/nachrichten
	Potsdam Institute for Climate Impact Research	http://www.pik-potsdam.de/aktuelles?set_language=de
	WWF Germany	http://www.wwf.de/themen-projekte/klima-energie
Skeptics	Analyse+Aktion	http://astrologieklassisch.wordpress.com/tag/klimawandel
	EIKE - Europäisches Institut für Klima und Energie	http://www.eike-klima-energie.eu
	Klimaüberraschung	http://www.klima-ueberraschung.de
	Klimaskeptiker	http://www.klimaskeptiker.info
United Kingdom		
Advocates	Greenpeace UK	http://www.greenpeace.org.uk/climate
	Oxfam UK	http://www.oxfam.org.uk/what-we-do/issues-we-work-on/climate-change
	Friends of the Earth UK	http://www.foe.org/projects/climate-and-energy
	WWF UK	http://www.wwf.org.uk/what_we_do/tackling_climate_change
Skeptics	The Global Warming Policy Foundation	http://thegwpcf.org
	Global Warming Hysteria	http://www.globalwarminghysteria.com/
	Climate Resistance	http://www.climate-resistance.org/
	Repealtheact	http://repealtheact.org.uk/
United States		
Advocates	Climate Central	http://www.climatecentral.org
	Greenpeace USA	http://www.greenpeace.org/usa/en/campaigns/global-warming-and-energy
	Worldwatch Institute	http://www.worldwatch.org/climate-energy
	WWF U.S.	http://www.worldwildlife.org/climate/index.html
Skeptics	The Heartland Institute	http://heartland.org/issues/environment
	Climate Depot	http://www.climatedepot.com
	C3 Headlines	http://www.c3headlines.com
	Watts Up With That?	http://wattsupwiththat.com

Table 5. Starting points for the snowball-sampling of websites in Germany, the United Kingdom, and the United States.

Appendix B: ANOVA Results

ANOVA results for the comparison of the expected topic probabilities over all actors by country.

Topic	F(1,813)	p
Causes / effects of climate change	8.895	0.003 **
Climate change evidence	20.003	< 0.001 ***
Climate change research	17.994	< 0.001 ***
Climate modeling	5.455	0.020 *
Climate politics	89.937	< 0.001 ***
Climate politics and science	3.663	0.056
Climate scepticism	24.673	< 0.001 ***
Doubting climate research	2.093	0.148
Economy and climate politics	18.649	< 0.001 ***
Energy consumption	27.207	< 0.001 ***
Energy sector	14.213	< 0.001 ***
Environmental activism	34.808	< 0.001 ***
Extreme weather	34.937	< 0.001 ***
Food / health	12.698	< 0.001 ***
Humanity	24.584	< 0.001 ***
Melting ice	9.949	0.002 **
Science	4.301	0.038 *
Scientific results / concensus	16.189	< 0.001 ***
Temperature	13.234	< 0.001 ***
U.S. environmental politics	17.94	< 0.001 ***
U.S. fiscal policy	1.721	0.190
Wildlife protection	0.051	0.823

Table 6. ANOVA results.

Appendix C: Fisher's Exact Test Results

Fisher's exact test results for every topic as top topic by country.

Topic	p
Causes / effects of climate change	< 0.001 ***
Climate change evidence	0.029 *
Climate change research	< 0.001 ***
Climate modeling	0.463
Climate politics	< 0.001 ***
Climate politics and science	0.401
Climate scepticism	1
Doubting climate research	0.013 *
Economy and climate politics	0.232
Energy consumption	0.005 **
Energy sector	0.008 **
Environmental activism	< 0.001 ***
Extreme weather	0.003 **
Food / health	0.321
Humanity	0.144
Melting ice	0.044 *
Science	0.819
Scientific results / consensus	< 0.001 ***
Temperature	< 0.001 ***
U.S. environmental politics	< 0.001 ***
U.S. fiscal policy	0.153
Wildlife protection	0.069

Table 7. Fisher's Exact Test results.