Multi-National Topics Maps for Parliamentary Debate Analysis

Markus Schaal
Berlin School of Economics
and Law
markus.schaal@hwr-berlin.de

Enno Davis
Berlin School of Economics
and Law
ennod@acm.org

Roland M. Mueller
Berlin School of Economics and Law
University of Twente
roland.mueller@hwr-berlin.de

Abstract

In recent years, automated political text processing became an indispensable requirement for providing automatic access to political debate. During the Covid-19 worldwide pandemic, this need became visible not only in social sciences but also in public opinion. We provide a path to operationalize this need in a multi-lingual topic-oriented manner.

Using a publicly available data set consisting of parliamentary speeches, we create a novel process pipeline to identify a good reference model and to link national topics to the cross-national topics. We use design science research to create this process pipeline as an artifact.

1. Introduction

The functioning of democracy relies on its public visibility, and therefore, the press sector is often seen as the fourth pillar of democracy. In times of social media and fake news, the fourth pillar is under attack and new means of transparency are needed. Beyond the scientific call for using text as data (cf. Grimmer and Stewart [1]) the automatic analysis of political debate is a must in the twenty-first century. It can help to re-create the long-lost transparency of democracy which is shadowed by fake news and social media bombs. We envision the emergence of publicly available apps such as a multi-national political debate browser that allows someone to navigate among auto-detected topics, and to find related documents and topics internationally without the need to manually edit topics or taxonomies, solely relying on automated topic analysis.

One ingredient for such an app would be a method to create multi-national topic models for the navigation and analysis of parliamentary debate. Probabilistic topic modeling methods are well-established for the creation of topic models in a single language. Political debate in different countries usually comes along with a different culture of discourse leading to different topics

and even incompatible term distribution within topics. Established cross-lingual topic modeling techniques do not suffice to solve the problem.

We therefore develop a novel process pipeline to identify cross-national reference topics and to link national topics to these cross-national topics. Following a design science research, we define a concrete instance of design science guidelines to create this process pipeline as an artifact. We use topic-specific dictionaries to align terms from various languages with their English translations before applying Latent Dirichlet Allocation (LDA) to the combined corpus. This is similar to other Multi-Lingual Probabilistic Topic Models (MuPTM) [2], which are based on document alignment, prior word matchings, or bilingual dictionaries. We chose LDA for its simplicity and availability of well-established open-source implementations.

After related work (Section 2) and methodology (Section 3), we develop and describe this novel process pipeline in Section 4. In short, we create a national topic model per corpus, translate top terms per topic, and apply LDA to the combined corpus with unified terms. The resulting multi-lingual topic model is our reference corpus for successful linking of related topics from various national corpora. Limitations and future work are mentioned in Section 5.

2. Related research

Text mining, natural language processing, and in particular Probabilisitc Topic Modeling (PTM) has received a growing research interest in the past decade. Latent Dirichlet Allocation introduced by Blei et al. [3] in their landmark paper is the current state-of-the-art method for topic modeling.

Various approaches exist for multi-lingual settings, see Vulić et al. [2] for a good overview. They define Multilingual probabilistic topic models (MuPTM) as a variation of monolingual probabilistic topic models (PTM) where topics are cross-lingual or language independent and term distributions per-topic are found
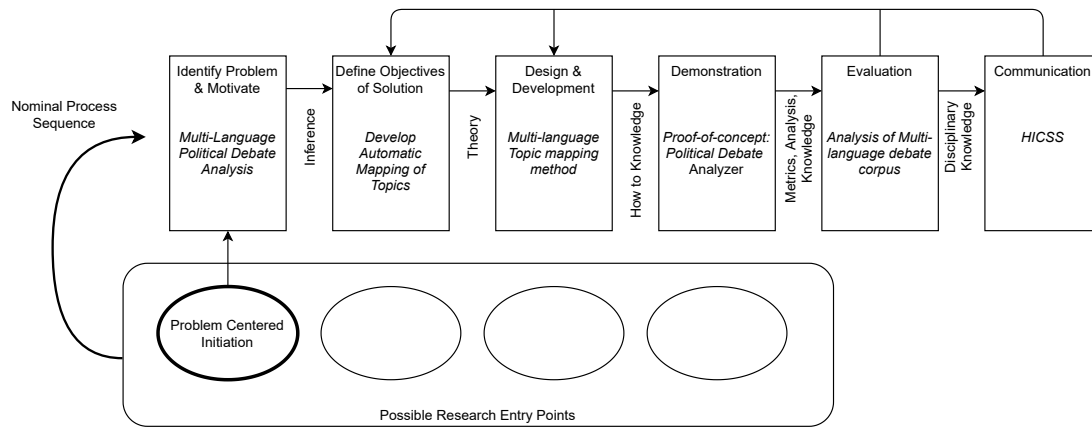
HǂCSS

Figure 1. DSRM Process Model

for each language independently. This is reached by aligning similar documents, sentences, or terms in different languages prior to modeling. De Smet and Moens [4] use aligned documents in Dutch and English from Wikipedia to classify events with an extended LDA approach in a multi-lingual setting. Zhao and Xing [5] use alignment at the sentence level to improve translation quality employing a probabilistic topic model. The JointLDA method of Jagarlamudi and Daumé [6] and also the MuTo approach of Boyd-Graber and Blei [7] merge aligned terms from different languages into single concepts before applying LDA. The latter method does not rely on document or sentence alignment and therefore applies to our scenario of unaligned parliamentary speeches from different countries.

3. Methodology

This paper uses design science research [8] as the main method to construct a design artifact for multi-national and multi-language political debate analysis. Peffers et al. [9] suggested a design science research methodology (DSRM) process model. We followed this DSRM for designing our artifact (see Figure 1). This design science research project has a problem-centered entry point. We identified as a real world problem the insufficient possibilities of multi-national and multi-language analysis of parliamentary debates. In order to help with these analyses, our design objective is to enable automatic mapping of political topics in a multi-language corpus. To achieve this objective, we designed a multi-language topic mapping method. We demonstrated the feasibility with a proof-of-concept prototype "Political Debate Analyzer". The method and prototype was analyzed based on a multi-language corpus of parliamentary

debates.

Hevner et al. [8] propose seven guidelines for design science research. In the following we present these guidelines and describe how we used them for creating our artifact:

Guideline 1: *Design as an Artifact.* We contribute to the design knowledge by suggestion of two new methods and test efficacy of the suggestions with an instantiation for multi-language political debate analysis. We suggest (i) a method for automatic stop word creation in the context of multi-language LDA and (ii) a method for matching multi-language LDA topics.

Guideline 2: *Problem Relevance.* Political debates are the cornerstone of every representative democracy. Cross-national political analysis; i.e., the comparison of political measures and societal behaviour across multiple nations can be enabled if topics can be aligned beyond language.

Guideline 3: *Design Evaluation.* Coherence scores are used to evaluate the quality of the cross-national (and cross-lingual) topic alignment. Implicitly, the reached coherence score is also a measure of the quality of the design.

Guideline 4: *Research Contributions.* This paper contributes to the LDA literature by addressing the challenges of multi-language political debate analysis.

Guideline 5: *Research Rigor.* While exploring the space of possible processing pipelines, the chosen coherence score measure is used to justify each design decision. Even though intuition sometimes is needed to find the direction, alternatives

are eliminated if they cannot be grounded on either coherence score optimization or rationally justified simplicity.

Guideline 6: *Design as a Search Process.* Different models and parameters are experimentally evaluated.

Guideline 7: *Communication of Research.* Detailed information of the used architecture is described in this paper.

4. Multi-national Political Topic Modeling

Topic models cluster unlabelled documents so that particular sets of words appear more likely in a single cluster, with the most prominent words becoming human-readable representations of the topic of each cluster. Latent Dirichlet Allocation [3] is a widely used modeling technique to achieve that goal. While there is no guarantee that the result is well interpretable, coherence measures have been proposed to distinguish between good and bad topics and average coherence scores across all topics are employed for comparing the overall quality of a model. In this work, we assume that we can use coherence score as a comparative measure to find our best processing pipeline for multi-national political topic modeling.

Following our design science guidelines, we started out with a problem-centered initiation for the design of our artifact, the process pipeline to create a multi-national and multi-lingual political topic model across our three national corpora of parliamentary debate, by application of probabilistic topic modeling. During the research, the following sub-goals emerged for the application of research rigor:

1. Treat the stop word problem fairly across all corpora, knowing that existing stop word lists are biased towards the language from which they were derived.

2. Find relevant terms and align them across all languages represented in the national parliamentary corpora, not assuming that sufficient shared discourse exists to rely on named entities.

3. After the application of probabilistic topic modeling, specifically LDA, to the multi-national corpus with cross-lingual matches, define how we can connect relevant topics across national corpora, ignore less relevant topics, and distinguish between multi-national and national topics.

4.1. The ParlSpeech V2 Data Set

column name	description	type
date	**speech date**	date
agenda	agenda item	str
speechnumber	order number	int
speaker	speaker name	str
party	party code	str
party.facts.id	party reference	int
chair	**speaker or chair**	bool
terms	term count in speech	int
text	**speech**	str
parliament	parliament name	str
iso3country	country code	str

Table 1. Column Definition. Adapted from Rauh and Schwalbach [10]

In recent years, more and more parliaments have provided API access to their debate archive. However, many differences exist in the concrete data format. We therefore rely on prior work of Rauh and Schwalbach [10] who prepared two versions of a corpus of parliamentary speeches, the so-called ParlSpeech (V2) data set. For reasons of simplicity, we limited ourselves to three nations: Germany, United Kingdom, and Spain. Two languages would not have been enough to derive patterns for multi-language application, while four or more languages would have added unneeded complexity. We chose the three nations to represent different language spaces but of course other language domains may add additional complexity. Each corpus is stored as a relational table with the columns shown in Table 1.

Initial investigations showed that even though the parliamentary corpora were stored in a unified manner, some corpora contained lots of comments (e.g., Germany) while other corpora had comments stored in separate rows (e.g., United Kingdom). This is due to the fact that the style of parliamentary debate differs a lot among countries. Here we had to make a difficult decision and removed all comments from all corpora in order to have single-speaker documents. Another issue was the fact that many data rows were marked as chair-contributions, usually not strongly relating to the content of the debate and we decided to remove them as well.

Finally, we decided to investigate the same time span for all three corpora which was 1996-03-27 to 2018-12-14, spanning 22 years in total. The resulting document counts per national corpus are given in parentheses: Germany (167943), Spain (108214), United Kingdom (1381804).
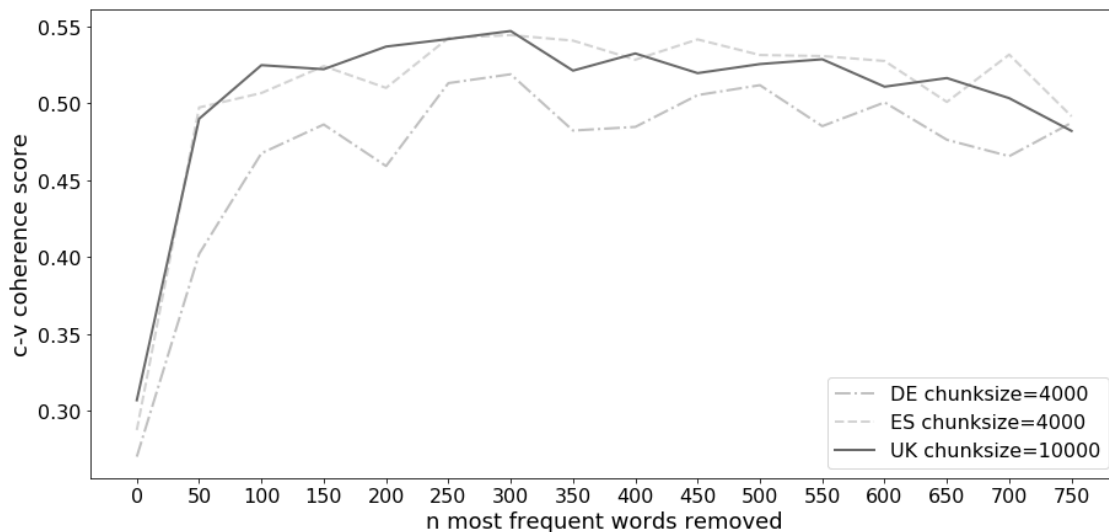
Figure 2. Coherence Score after LDA with Top n most frequent terms removed (0-750, 50-step)

4.2. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation is a probabilistic topic modeling technique, that co-clusters documents and terms in such a way that certain term clusters appear more likely in a particular doc-cluster than in all other doc-clusters. The result of LDA therefore is a matrix of probabilities per topic, term and document. The most prominent terms per topic-cluster can be viewed as the most relevant terms for that topic and function as a machine-generated but human-readable representation for each topic.

LDA models are created by an iterative procedure to find the best possible clusterings and probability distribution. Various libraries exist that implement LDA. We chose Gensim as a good and widely used library by Řehůřek and Sojka [11] that implement a number of variants of LDA and other probabilistic topic modeling techniques. LDA is controlled by a number of parameters, most of which we left at default. Topic num, chunksize, and passes are the most important decisions to take. In each iteration chunksize documents are processed until the corpus is exhausted. Topic num controls the number of topics. If more than one pass is chosen, the corpus is processed multiple times.

4.3. Coherence Score as Evaluation Method

Various coherence scores have been introduced as a formal measure for the quality of a topic model. Röder et al. [12] introduced **c-v** as a novel coherence measure which they found to have the highest correlation with human ratings. Assuming that the algorithm is fundamentally correct and the word lists have been perfectly pre-processed, **c-v** will be our method to evaluate particular topics chosen, model quality, and even a guidance for design decisions based on model or topic quality.

The following patterns were used to employ coherence score as a means towards design choices:

1. In cases of optimization across a whole corpus, average coherence score among all topics created by probabilistic topic modeling would determine which model to choose. This was the case for Section 4.5 Pass Ia: Frequency Cap Optimization and Section 4.6 Pass Ib: Topic Number Optimization.

2. In situations where design choices had no output that could be immediately measured in terms of coherence score, we explored various alternatives and tried to check the coherence scores of the linkable topics in the end result, where average coherence score of linked topics was the ultimate measure. Obviously there are millions of further design choices between the choice in question and the final artifact so that in these cases we can only argue indirectly about the concrete reasons for the choice. As a baseline, we decided that the minimum coherence score to be accepted should be 0.5 and used as our iron[1] standard.

---

[1] A best (or human-made) standard is usually called a gold standard.

4.4. Pre-Processing

Each corpus consists of a set of documents. Before we can process them in a quantitative way, we had to remove punctuations and numbers, i.e., digits 0-9. Afterwards, in order to process words irrespective of their grammatical form, we applied the spaCy Lemmatizer, an industry standard meta-lemmatizer, that uses best available practices per language. Finally, we removed single-letter literals and made everything lowercase.

Design Choice 1: Simple pre-processing as described.

Decision Method: Rationally justified simplicity.

4.5. Pass Ia: Frequency Cap Optimization

Stop words (i.e., words that hinder or do not contribute to quantitative processing of language) are to be removed. Previous research has suggested various stop word lists for various purposes and languages. However, in order to be fair, we omit hand-crafted stop word lists and try to find a fair boundary for frequency-based removal by measuring coherence scores after LDA as a function of parameter n for top-n frequent term removal.

For all three national corpora, after pre-processing, we removed the N most frequent words with the *filter_n_most_frequent(remove_n)* method of the Gensim [11] Dictionary object, where frequency is the number of documents in which a term occurs. Then we applied LDA with a standard of 50 topics, chunksize appropriate to ensure that more than 20 iterations occur in a single pass, and then we evaluated the resulting **c-v** coherence score. Results are shown in Figure 2.

Design Choice 2: Use 50 topics for Top N most frequent word optimization.

Decision Method: Rationally justified simplicity.

By applying the **c-v** coherence measure and using the maximum (average) coherence per national corpus, we applied our design science guideline to decide for a particular alternative based on our universal quality measure which is the **c-v** coherence measure as defined by Röder et al. [12].

The absolute maximum for all three national corpora was at 300, so we created a frequency-based stop-word list for all three corpora at this boundary, respectively and applied this stop word list for all sub-sequent processing.

Design Choice 3: Remove 300 most frequent words from all national corpora.

Decision Method: Best choice for model coherence score **c-v**.

4.6. Pass Ib: Topic Number Optimization

id	topic terms	c-v
7 EN	energie, klimaschutz, energiewende *energy, climate protection, energy transition*	0.738
15 EN	pflege, arzt, versorgung *care, doctor, supply*	0.736
6 EN	bundeswehr, einsatz, soldat *armed forces, use, soldier*	0.719
11 EN	türkei, syrien, menschenrechte *turkey, syria, human rights*	0.709
19 EN	projekt, infrastruktur, straße *project, infrastructure, street*	0.665

Table 2. Top-5 (out of 25) topics for Germany

id	topic terms	c-v
3 EN	fiscal, imponer, impuesto *prosecutor, impose, tax*	0.752
0 EN	educativo, educación, formación *educational, education, training*	0.681
8 EN	aguar, andalucía, valenciano *water, andalusia, valencian*	0.673
11 EN	crecimiento, crisis, déficit *growth, crisis, deficit*	0.661
9 EN	justicia, judicial, civil *justice, judicial, civil*	0.652

Table 3. Top-5 (out of 30) topics for Spain

id	topic terms	c-v
15	defence, armed, war	0.776
5	crime, prison, victim	0.731
0	care, nhs, hospital	0.727
6	international, security, foreign	0.725
33	rail, transport, train	0.721

Table 4. Top-5 (out of 35) topics for United Kingdom

Average coherence score does not only differ for different Top N removed words, but also differs depending on the amount of topics. Assuming that the optimal number of topics might be different for different languages and national corpora, we did a comparison of average coherence score for different topic numbers as input to the LDA algorithm. Obviously we removed Top 300 Words as decided in Section 4.5. The result is shown in Figure 3.
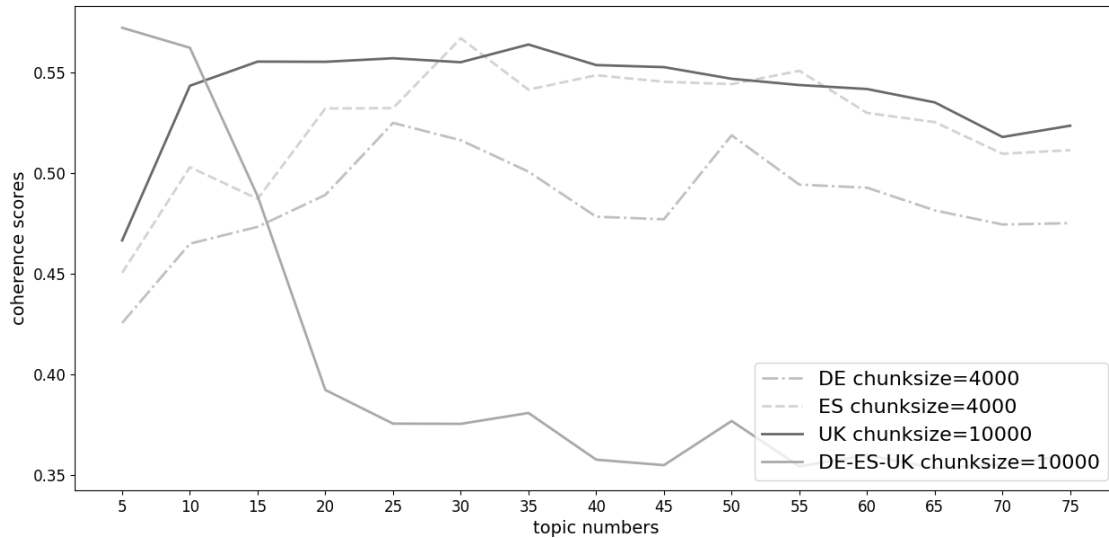
Figure 3. Coherence Score after LDA (removing 300 words) with k topics (5-75, 5-step)

The maximum coherence score for Germany was at 25 topics, for Spain at 30 topics and for the United Kingdom at 35 topics. Accordingly, we used the respective topics models for cross-national further processing. Note, we also show the coherence score for the combined corpus already here in this figure. We will come back to this at some later point.

Design Choice 4: Choose 25 topics for Germany, 30 topics for Spain, and 35 topics for United Kingdom, and more generally apply automatic selection of topic num for each national corpus.

Decision Method: Best choice for model coherence score **c-v**.

For the chosen topic numbers per corpus, the five most coherent topics are represented by their top terms in Tables 2, 3, and 4. For Germany and Spain, the second row for each topic is the English translation, marked by *EN*.

4.7. Pass II: Unified Multi-National Corpus

Each national parliament has a unique culture of debate and domain of discourse, in addition to its respective language. Therefore, a topic-model built on top of a union text corpus would not reflect any of the national parliament's discourses correctly.

We therefore suggest the creation of a specifically designed corpus with important terms (per national corpus) linked across languages. This corpus is also trained with LDA and used as a reference corpus to create cross-national topic links. The method would also distinguish between national and multi-national topics by measuring the weighted connection between the topics.

For our multilingual setting, we followed a similar approach as Boyd-Graber and Blei [7], but we are looking at three languages here and aim to serve as many languages as needed. Therefore, and in order to avoid the combinatorial explosion of multi-language ambiguities, we decided to use the English language as an anchor for any bi-lateral translations (i.e., we omit dictionaries between non-English languages).

Design Choice 5: For alignment across languages, we create hard links between any two words that can be found in an electronic dictionary, but we link only between English and non-English.

Decision Method: Rationally justified simplicity.

For each non-English corpus, we translate the most prominent terms in each topic into English and replace both sides of the translation by its unified id in the respective corpus. Each single word is translated by Google translate, with source and target language specified. Remember also that the words are already lemmatized.

Design Choice 6: Only translate the most prominent terms in the national LDA model.

Decision Method: Rationally justified simplicity.

Here is the concrete algorithm:

1. For top 20 words in each topic, we create the translations (with Google) from each non-English corpus into English. To keep a measure of importance per token we sum the LDA membership probabilities per topic.

2. For each translation we assign an ID.

3. If multiple translations have the same target word: Keep only the most important translation to be decided by the importance of the source term in the topics of the source language.

4. Translations of different languages in English that target the same word in English are unified, i.e., they receive the same ID.

5. For multi-term English translations, those are considered as multi-token representations.

6. Filter the translations with the stop words from Section 4.5.

7. Finally, replace each translated word or multi-term in each corpus by their unified ID

8. Create the multi-national corpus as a combination of the individual corpora.

Finally we applied LDA and also measured the resulting coherence score for 5-75 topics in 5-steps. The results are also in Figure 3. Note, the coherence score for the unified corpus drops dramatically from 0.562 at 10 topics to 0.392 at 20 topics. At the same time, the individual national corpora increase their coherence score towards the optimum. This seems to suggest that it becomes harder to create good unified topics across all corpora while the topics of the individual corpora become better. This is supported by our exploratory experiments. Even at e.g., 15 topics, we found that very different topics from different national corpora where combined into single topics with high coherence in the unified corpus. One topic e.g., contained both pension- and energy-related terms, which would lead to confusion in the linking phase 4.8.

We do not know yet how many topics from the national corpora can be linked to a corresponding reference topic and we do not know how many of those topics only relate to the respective country. In the range between 20 topics and 35 topics (the maximum number of topics in any national topic model), we therefore chose the minimum. Also, during our exploratory experiments, we did not find topics in our unified model that combined incompatible topics from different national corpora when using 20 topics or more.

Design Choice 7: Choose 20 as topic number for the reference corpus.

Decision Method: Rationally justified simplicity and the need to have topics in the reference corpus that do not combine multiple topics.

4.8.   Linking the Topics

For each national corpus, we link topics to the reference topic model as follows:

1. We only consider national topics with a minimal coherence score of 0.5.

2. We calculate cosine similarity between all pairs of topics from the two different topic models, i.e., between any pair of one topic from the national model and one topic from the reference model.

3. We sort these pairs by descending similarity and cut off at 0.1 similarity.

4. Each highest pair between two topics in this sorted list where neither side has been linked by a higher pair becomes a link.

We choose 0.5 as minimum coherence because it is not feasible to try to link weaker topics that may have a higher cosine similarity than better topics. Also, we choose 0.1 as a threshold for cosine similarity because higher values would have removed too many links and lower thresholds would have led to random linking.

Design Choice 8: Choose 0.5, the iron standard, as minimum coherence.

Decision Method: Rationally justified simplicity.

Design Choice 9: Choose 0.1 as minimum cosine similarity.

Decision Method: Rationally justified simplicity.

4.9.   Results and Discussion

Applying our method, we receive the result in Table 5. The two terms marked between stars (i.e., *progress* and *white*) indicate words from the original term list (i.e., they were never translated). Note, not all reference topics are mapped but only 11 out of 20. The average coherence score of unmapped reference topics is 0.442 and the average of mapped reference topics is 0.373. It seems that reference topics with a lower coherence are much more likely to act as a glue between topics from different national corpora.
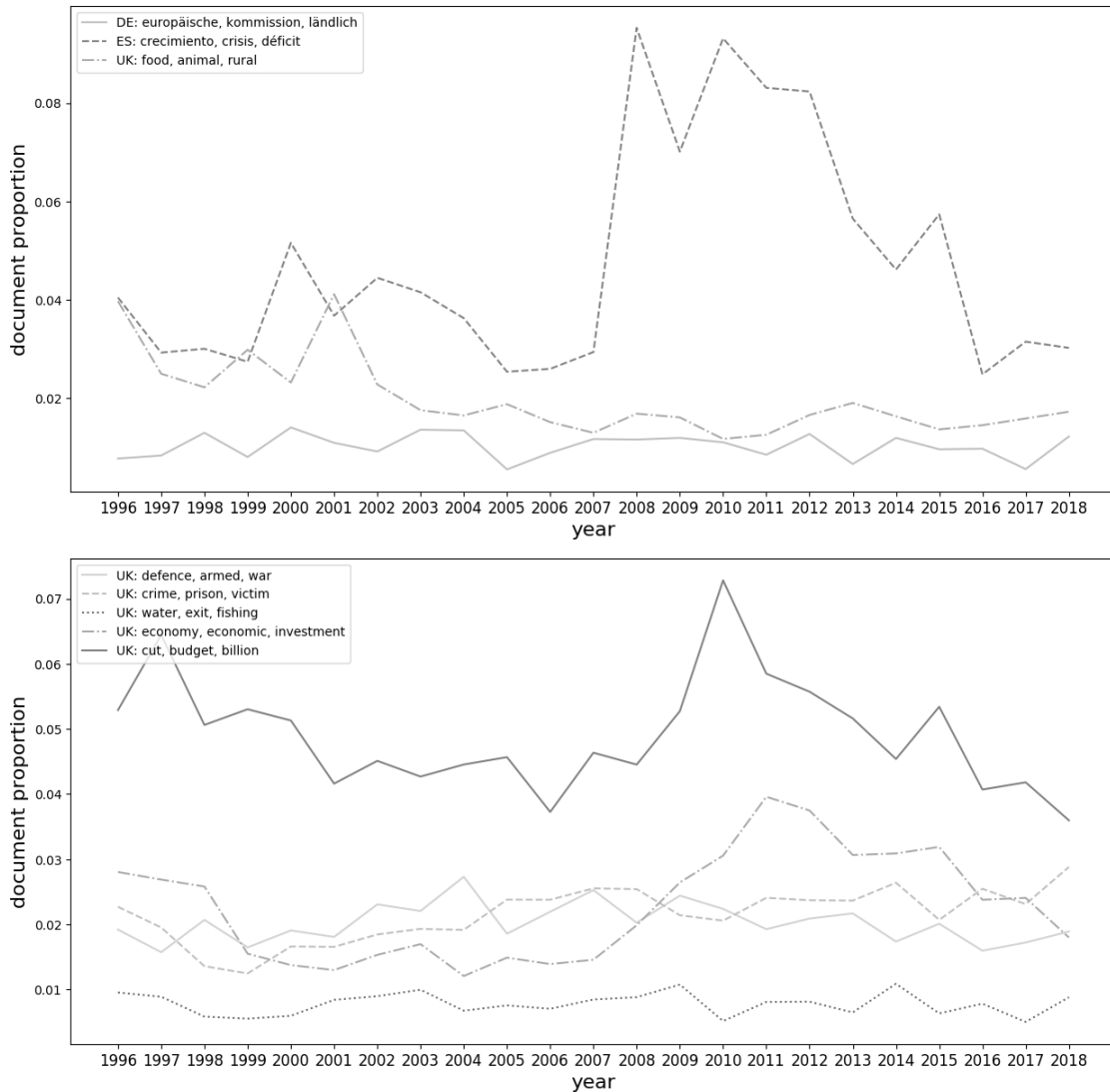
Figure 4. Development of topics across time

The validity of our decision to use a reference model with low average coherence score is illustrated here by topic number 9 with a coherence score of 0.514. This topic is neither mapped to all three countries nor is it very concise, i.e., it apparently spans a broad subject range from *corruption, democracy, democratic* (Spain) to *vote, referendum, election* (United Kingdom).

In Tables 6, 7, and 8, we present the top 5 unlinked topics for each of the three countries. For Germany and Spain, topic terms are also translated into English. Topics with high score that could still not be linked are topics of national interest only. For example see topic *water, andalusia, valencian* or *catalonia, catalan, constitution* which are Spain-only topics.

Finally we demonstrate one possible use of such multi-national topic maps, that is the comparison of topics across time, either by country (for cross-national topics) or by topic, see Figure 4. The cross-national topic shows the evolution over time for topic number 6 in Table 5. Here we see the effect of the 2008-2014 crisis in Spain.

All of our design choices are either independent of the number of national corpora (Design Choices 1, 2, 5, 6, 8, and 9) or it can be safely assumed that only minor changes (Design Choices 3, 4, and 7) will occur with additional corpora. Therefore we assume that this research easily extends to additional national corpora and languages.

#	Germany	Spain	United Kingdom	Reference	c-v
1	europa, europäische, eu *europe, european, eu*	internacional, humano, inmigración *international, human, immigration*	international, security, foreign	international, europe, fight	0.345
2	zahlen , steuer, haushalt *pay, tax, budget*	fiscal, imponer, impuesto *prosecutor, impose, tax*	tax, credit, universal	tax, bank, investment	0.357
3	-	mujer, igualdad, familia *woman, equality, family*	young, woman, immigration	violence, language, young	0.387
4	bildung, forschung, ausbildung *education, research, training*	educativo, educación, formación *educational, education, training*	school, education, pension	education, school, university	0.394
5	verbraucher, produkt, is *consumer, product, is*	financiero, entidad, información *financial, entity, information*	legislation, clause, regulation	information, protection, control	0.302
6	europäische, kommission, ländlich *european, commission, rural*	crecimiento, crisis, déficit *growth, crisis, deficit*	food, animal, rural	crisis, rural, agriculture	0.344
7	opfer, schutz, gewalt *victim, protection, violence*	justicia, judicial, civil *justice, judicial, civil*	law, court, legal	justice, court, civil	0.371
8	-	trabajador, laboral, formación *worker, labor, training*	company, market, consumer	worker, contract, employer	0.36
9	-	corrupción, democracia, democrático *corruption, democracy, democratic*	vote, referendum, election	democracy, *progress*, *white*	0.514
10	pflege, arzt, versorgung *care, doctor, supply*	sanitario, sanidad, salud *health, health, health*	care, nhs, hospital	care, professional, quality	0.388
11	projekt, infrastruktur, straße *project, infrastructure, street*	inversión, sostenibilidad, obrar *investment, sustainability, work*	rail, transport, train	investment, infrastructure, energy	0.344

Table 5. Multi-National Topic Map

5. Conclusions

Following the design science principles, we created an artifact, our pipeline to process multi-national parliamentary speech corpora for augmenting cross-national analysis and topic navigation, with the help of our proposed multi-national topic models.

In particular, the following three research contributions are made. First, we show a data driven and general approach for filtering corpora-specific stop words with LDA. It is possible to filter redundant words without any pre-defined stop word list for topic modelling based on a word frequency approach.

Second, we provide a method to join the corpora per country into a single corpus and to use probabilistic topic modelling on this single corpus to create a reference topic model for cross-national topic linking. We also propose a method for this kind of topic linking.

Germany-Only	c-v
energie, klimaschutz, energiewende *energy, climate protection, energy transition*	0.738
bundeswehr, einsatz, soldat *armed forces, use, soldier*	0.719
umwelt, grenzwert, nachhaltig *environment, limit value, sustainable*	0.612
international, afghanistan, soldatinnen *international, afghanistan, soldiers*	0.52
kind, flüchtling, familie *child, refugee, family*	0.482

Table 6. Germany-only Topics

Spain-Only	c-v
aguar, andalucía, valenciano *water, andalusia, valencian*	0.673
penal, delito, código *criminal, crime, code*	0.649
vivienda, joven, alquiler *housing, young, rent*	0.648
energético, energía, eléctrico *energetic, energy, electric*	0.64
cataluña, catalán, constitución *catalonia, catalan, constitution*	0.596

Table 7. Spain-only Topics

United Kingdom-Only	c-v
defence, armed, war	0.777
crime, prison, victim	0.731
water, exit, fishing	0.662
economy, economic, investment	0.629
cut, budget, billion	0.587

Table 8. United Kingdom Topics

Based on the linked topics and the coherence scores of national and reference topics, we can find topics that are either shared in a cross-national discourse, are country-specific, or are simply too broad or undefined to be relevant for cross-national topic navigation.

Finally, we demonstrated the use of a topic model created with our artifact to enable the cross-national analysis of topical discourse over time in different parliaments.

One of the limitations of our work is the lack of comparison with other methods for creating cross-lingual political transparency. We also did not optimize the hyper-parameters of LDA. Future work should also increase the applicability by design of good user-interfaces for multi-national topic navigation, potentially not only in parliamentary speeches from the ParlSpeech data set but also in publicly available related news and social media posts. Stream-processing and online learning in real-time would be another future direction to allow for the navigation in current political information.

References

[1] J. Grimmer and B. M. Stewart, "Text as data: The promise and pitfalls of automatic content analysis methods for political texts," *Political Analysis*, vol. 21, no. 3, p. 267–297, 2013.

[2] I. Vulić, W. De Smet, J. Tang, and M.-F. Moens, "Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications," *Information Processing & Management*, vol. 51, p. 111–147, Jan 2015.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, p. 993–1022, Mar. 2003.

[4] W. De Smet and M.-F. Moens, "Cross-language linking of news stories on the web using interlingual topic modelling," in *Proceedings of the 2nd ACM workshop on Social web search and mining*, pp. 57–64, 2009.

[5] B. Zhao and E. P. Xing, "BiTAM: Bilingual topic AdMixture models for word alignment," in *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, (Sydney, Australia), pp. 969–976, Association for Computational Linguistics, July 2006.

[6] J. Jagarlamudi and H. Daumé, "Extracting multilingual topics from unaligned comparable corpora," in *Advances in Information Retrieval* (C. Gurrin, Y. He, G. Kazai, U. Kruschwitz, S. Little, T. Roelleke, S. Rüger, and K. van Rijsbergen, eds.), pp. 444–456, Springer, Berlin, 2010.

[7] J. Boyd-Graber and D. M. Blei, "Multilingual topic models for unaligned text," in *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, (Arlington, Virginia, USA), p. 75–82, AUAI Press, 2009.

[8] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, p. 75–105, 2004.

[9] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, p. 45–77, Dec 2007.

[10] C. Rauh and J. Schwalbach, "The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies," 2020. https://doi.org/10.7910/DVN/L4OAKN.

[11] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, (Valletta, Malta), pp. 45–50, ELRA, May 2010.

[12] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, (New York, NY, USA), p. 399–408, Association for Computing Machinery, 2015.