USING NATURAL LANGUAGE PROCESSING METHODS TO ANALYZE THE EVOLUTION OF PARTY POSITIONS IN THE GERMAN BUNDESTAG

ABSTRACT

Migration has always been a politically provocative and emotionally charged issue. At the same time, it is one of the most important core political issues in recent history. This project examines how the debate in the German Bundestag has been influenced by the parties' normative positions on migration. A mixed-methods approach was used in this study. In order to get an overview of the patterns in the debates, computational methods such as term frequency analysis, LDA topic modeling and keyword embedding analysis are used. The methodological extension of historical source analysis in the digital humanities is practically demonstrated in this project.

DATA BASE

The project used the Open Discourse corpus data. The corpus consists of five main tables: “speeches”, “factions”, “politicians”, “electoral term”, “contributions simplified” and “contributions extended”. “speeches” table contains all speeches held in parliament with meta information. 907644 observations and 12 variables.
For more information see: Open Discourse Data Dump

1 Preprocessing: lemmatization, remove punctuation and numbers

2 Applying LDA Topic Model with MALLET

3 Testing semantics for selected keywords with Word2Vec

4 Frequency analysis of selected keywords

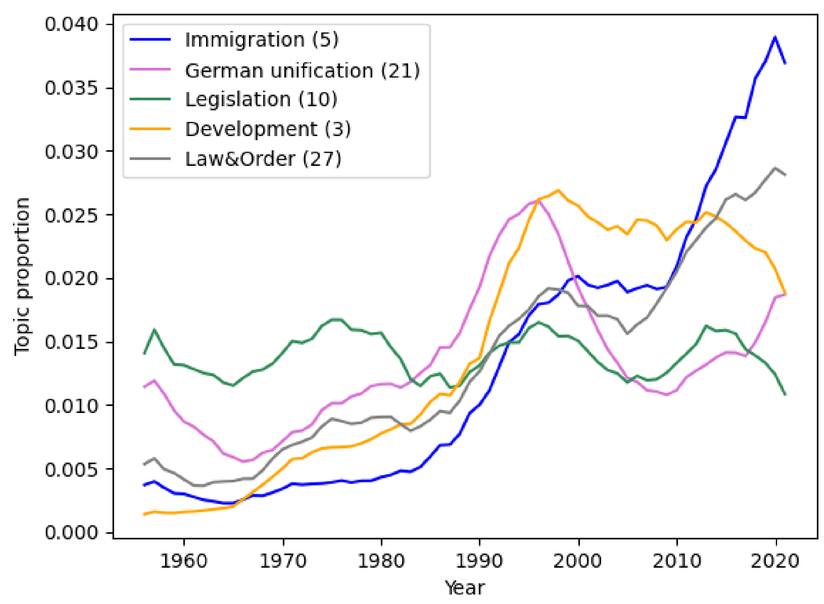
5 Topic distribution over time, keyword embedding in topics, topic-related speech filtering

6 Historical analysis of the speeches

CONCLUSION

This analysis highlights the advantages of a mixed-methods approach. It showed that migration debates in the Bundestag vary in importance through quantitative analysis. In addition, there was evidence that trends in party discourse could be approximated. The theoretically derived attitudes of the SPD and the CDU/CSU in the sources were confirmed in the first analysis. In line with previous historical research, topic modeling made it possible to visualize thematic connections in the speech content. This method opens up many possibilities for analyzing the corpus. In this study, the focus was on speeches in which the topic of migration was the most important issue. This approach proved feasible given the size of the corpus.

Output of topic distribution over time



More examples of the visualization of the results can be found on the website of the project:
https://reebmaren.github.io/thesis-bundestag-debate/index.html

MAREN REEB

Subject: Digital Methodology in the Humanities and Cultural Studies faculty 07 - History and cultural sciences of the Johannes Gutenberg University Mainz