

Language- and political polarization of parliamentary speeches between 1998-2020

Keywords: Political Polarization, Natural Language Processing, Document Embedding, Deep Learning, BERT

Extended Abstract

Politicians have an influential role in the construction of social reality and the thematization of topical issues of a country's public sphere. They simultaneously have an effect on and are influenced by the political polarization of society. Therefore, studying the political speeches of the members of a parliament provides us with unique insights into this process and about society.

Parliamentary- and political communication in general has undergone significant change over the last two decades. The aim of this research was to examine language- and political polarization in Hungary through the speeches of the members of the Hungarian Parliament. Our work applied Natural Language Processing, Machine- and Deep Learning techniques on transcripts of the parliamentary speeches held by politicians of the Parliament from 1998 to 2020, roughly 400.000 unique texts.

After the basic preprocessing routines, we first used a time-dependent Structural Topic Model (Roberts et al., 2019) to identify the topics that distinguish particular groups – e.g. parliamentary factions and parties – of speakers in each parliamentary term (or year). This model identifies distinct topics of a corpus in an exploratory fashion based-on word usage, and assigns a word distribution to each topic and a topic distribution to each text. These topics are highly interpretable, but the relationship of topics to each other is hard to measure since semantic information is not used in the model.

To address this, we constructed Word Mover's document Embeddings (Wu et al., 2018) of speeches that allowed us to measure the similarities of the previously identified topics. Tracking the usage and similarities of topics helped reveal the political communication strategy of each party relative to others throughout the period.

Thirdly, based on the complete preprocessed transcripts of the speeches we built state-of-the-art classification models (XGBoost (Hastie et al., 2019) and BERT (Devlin et al., 2018)) to predict the party affiliation of the speakers. Evaluating these models for each party and parliamentary term (or year) enabled us to examine the changes in amplitude and characteristics of language- and political polarization over time.

Based on the results of the analysis, the patterns of change in political communication strategies were mapped both for the governing- and the opposition parties, and individual politicians who have been key figures throughout the examined period. Within-party use of topics showed a steady increase in homogeneity for right-wing parties, and a decrease for left-wing parties. Moreover, according to the results of the classification models, increasing language- and political polarization was observed between the governing parties and all other parties of the Hungarian Parliament.

The patterns of intra-party homogenization and inter-party polarization evident in the speeches were in line with the characteristics of neopopulism-theory (Csigó 2016), according to which a speculative mediatized political space emerged where politicians observe the strategies employed by other politicians to appear popular and distinguish themselves from others by creating self-referential communication bubbles.

References

- Csigó, P. (2016). *The Neopopular Bubble. Speculating on" the People" in Late Modern Democracy*. Central European University Press.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4), 1307-1340.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). STM: An R package for structural topic models. *Journal of Statistical Software*, 91, 1-40.
- Wu, L., Yen, I. E., Xu, K., Xu, F., Balakrishnan, A., Chen, P. Y., ... & Witbrock, M. J. (2018). Word Mover's Embedding: From word2vec to document embedding. *arXiv preprint arXiv:1811.01713*.