A Semantic Map of the European Parliament

Giorgi Gogelashvili [* 1] **Samia Haque** [* 2] **Jakob Kleine** [* 3] **Dennis Stroh** [* 4] **Quirin Unterguggenberger** [* 5]

Abstract

Motivated by the rise of populism in Europe since the late 1990s, this study investigates ideological shifts in European Parliament (EP) speeches using natural language processing. Drawing on the novel ParLawSpeech dataset (Schwalbach et al., 2025) which contains 574,199 speeches from 1999 to 2024 alongside metadata on speaker identity, we use sentence embedding models to examine the semantic content and emotional tone of parliamentary debates over time.

We expect that speech embeddings will form clusters reflecting party affiliation and ideological alignment. In step with recent political developments, we further hypothesize an increase in negative sentiment within the immigration debate among centrist and right-wing groups, accompanied by growing semantic similarity between these two factions over the past two decades. Finally, we test whether established migration-related narratives associated with right-wing populism can be identified in parliamentary discourse and how their prevalence has developed over time.

1. Introduction

The continued success of right-wing populist parties in the 21st century is widely regarded as a major threat to European democracy and integration (Fossum, 2023; Rummens, 2017). Populist rhetoric is commonly defined as constructing an antagonism between a 'pure people' and a 'corrupt elite' (Mudde, 2007). Right-wing populism is also closely tied to the issue of immigration. Parties of this ideology have played a central role in the increasing politicisation of immigration (Hutter & Kriesi, 2022), which represents a crucial factor for their political success (Kende & Krekó, 2020). Over the past decade, immigration has become an increasingly salient issue in European election campaigns (Dekeyser & Freedman, 2023) as well as in media coverage (Greussing & Boomgaarden, 2017).

Electoral gains of populist parties have manifested in significant changes of parliamentary discourse (Schwalbach, 2023). A recent quantitative analysis of EP speech embeddings has identified a gradual increase in emotional rhetoric since 1999, with right-wing populist groups leading the trend (Subtil & Verger, 2024). In the German national parliament, an LLM-based study has revealed increasing anti-solidarity messaging around immigration, not only for right-wing, but also christian-conservative and liberal parties (Kostikova et al., 2025). This trend begins around 2015, which marks the onset of the so-called 'refugee crisis' (Brücker et al., 2020).

More fine-grained analyses of the migration discourse have revealed the use of common underlying narratives, defined as 'selective depictions of reality' and 'patterns of interpretation' through which the issue is relayed to the public. Social media posts from populist leaders commonly employ anti-immigrant frames like 'immigrants take our jobs' or anti-establishment narratives such as 'our sovereignty is under threat' (Seiger et al., 2025).

2. Data and Methods

2.1. Dataset description

2.2. Data preprocessing

Handling Duplicates (TODO)

Translation To keep the speeches most comparable in the embedding space, we use English translations instead of the orignal speeches. Until the year X (TODO), the Parllaw dataset includes a machine translation for each speech. The remaining X (TODO) translations were created using Gemini (2.5-flash, TODO: add citation), a LLM capable of translating texts across various domains (TODO citation). We checked that Gemini *1)* did not re-formulate speeches that were already in English and that *2)* its translations are comparable to Parllaw's in the embedding space. For this, we tested its translations on a random sample of speeches that had already been translated by Parllaw. Gemini *1)* preserved speeches which were already in English [1] and *2)* created translations whose embeddings are very similar to Parllaw's (bootstrapped 0.95 confidence interval of mean

---

[1](TODO add quantifier) Since we created translations before extensively cleaning the dataset, some English speeches included bracketed language flags that led to Gemini re-translating the English speeches. These reformulations are however almost identical to the original speech. Therefore, we accepted those instances where Gemini failed to recognize English texts.

cosine similarity: 0.969, n=1001). Thus, we assume that Gemini's and Parllaw's translations are similar enough to conduct our analysis under the assumption that all translations stem from the same source after filling in the missing translations with Gemini's. However, we note that the mixture of two translation approaches might nevertheless introduce a bias to our dataset, that we have to check for. (TODO: did we check for that)

2.3. Methods

2.3.1. TOPIC MODELLING WITH LDA

To identify how parties talk about migration, we first have to assign speeches to a semantic topic. For this purpose, we use Latent Dirichlet Allocation *(LDA)* [TODO: source]. (TODO why LDA?) LDA is a probabilistic topic model. It assumes that in the analyzed corpus (here: the collection of speeches) there is a set of topics, which are probability distributions over all words in the corpus. It considers each document (here: a speech) as a bag of words that were sampled from these topics. For example, if a topic had high probabilities for the words "fish", "net", "water", then documents covering "fishing" would (under LDA's assumptions) have a high probability of being labelled as that topic.

We tested different parameters (numer of topics, number of iterations over the dataset) and compared the resulting topic coherence (TODO explanation) and the fidelity of the topics through manual inspection. Our final model contains 30 topics (for 10 iterations) — one of which assigns highest probabilities to the words X, Y, Z (TODO words), which we call "migration topic".

For each topic, the model assigns each speech in the corpus a probability of covering that topic. To find speeches covering migration, a minimum topic probability was identified manually: two authors rated a sample of 100 speeches whose migration score (i.e. the assigned probability of the LDA model for the migration topic of that speech) was in the range [X, Y] (TODO range) which is where we suspected the relevance threshold.

(TODO finish this)

References

Brücker, H., Kosyakova, Y., and Vallizadeh, E. Has there been a "refugee crisis"?: New insights on the recent refugee arrivals in germany and their integration prospects. *Soziale Welt*, 71(1/2):pp. 24–53, 2020. ISSN 00386073. URL https://www.jstor.org/stable/27004992.

Dekeyser, E. and Freedman, M. Elections, Party Rhetoric, and Public Attitudes Toward Immigration in Europe. *Political Behavior*, 45(1):197–209, March 2023. ISSN 0190-

9320, 1573-6687. doi: 10.1007/s11109-021-09695-w. URL https://link.springer.com/10.1007/s11109-021-09695-w.

Fossum, J. E. In What Sense Does Right-Wing Populism Pose a Democratic Challenge for the European Union? *Social & Legal Studies*, 32(6):930–952, December 2023. ISSN 0964-6639. doi: 10.1177/09646639231153306. URL https://doi.org/10.1177/09646639231153306.

Greussing, E. and Boomgaarden, H. G. Shifting the refugee narrative? An automated frame analysis of Europe's 2015 refugee crisis. *Journal of Ethnic and Migration Studies*, 43(11):1749–1774, August 2017. ISSN 1369-183X, 1469-9451. doi: 10.1080/1369183X.2017.1282813. URL https://www.tandfonline.com/doi/full/10.1080/1369183X.2017.1282813.

Hutter, S. and Kriesi, H. Politicising immigration in times of crisis. *Journal of Ethnic and Migration Studies*, 48(2):341–365, January 2022. ISSN 1369-183X, 1469-9451. doi: 10.1080/1369183X.2020.1853902. URL https://www.tandfonline.com/doi/full/10.1080/1369183X.2020.1853902.

Kende, A. and Krekó, P. Xenophobia, prejudice, and right-wing populism in East-Central Europe. *Current Opinion in Behavioral Sciences*, 34:29–33, August 2020. ISSN 2352-1546. doi: 10.1016/j.cobeha.2019.11.011. URL https://www.sciencedirect.com/science/article/pii/S2352154619301299.

Kostikova, A., Pütz, O., Eger, S., Sabelfeld, O., and Paassen, B. LLM Analysis of 150+ years of German Parliamentary Debates on Migration Reveals Shift from Post-War Solidarity to Anti-Solidarity in the Last Decade, September 2025. URL http://arxiv.org/abs/2509.07274. arXiv:2509.07274 [cs].

Mudde, C. *Populist Radical Right Parties in Europe*. Cambridge University Press, Cambridge, 2007. ISBN 978-0-521-85081-0. doi: 10.1017/CBO9780511492037. URL https://www.cambridge.org/core/books/populist-radical-right-parties-in-europe/244D86C50E6D1DC44C86C4D1D313F16D.

Rummens, S. Populism as a Threat to Liberal Democracy. In Kaltwasser, C. R., Taggart, P., Espejo, P. O., and Ostiguy, P. (eds.), *The Oxford Handbook of Populism*, pp. 0. Oxford University Press, October 2017. ISBN 978-0-19-880356-0. doi: 10.1093/oxfordhb/9780198803560.013.27. URL https://doi.org/10.1093/oxfordhb/9780198803560.013.27.

Schwalbach, J. Talking to the populist radical right: A comparative analysis of parliamentary debates. *Legislative Studies Quarterly*, 48(2):371–397, 2023.

Schwalbach, J., Hetzer, L., Proksch, S.-O., Rauh, C., and Sebők, M. Parllawspeech. (Version 1.0.0) [Data set]. GESIS, Cologne. https://doi.org/10.7802/2824, 2025.

Seiger, F., Kajander, N., Neidhardt, A.-H., Scharfbillig, M., Dražanová, L., Deuster, C., Krawczyk, M., Blasco, A., Icardi, R., Tzvetkova, M., Bakker, L., Olivo Rumpf, K., and European Commission (eds.). *Navigating migration narratives: research insights and strategies for effective communication*. Publications Office, Luxembourg, 2025. ISBN 978-92-68-28629-6. doi: 10.2760/2350572.

Subtil, H. and Verger, V. Emotional rhetoric and the rise of populism in the european parliament. Policy Brief 108, IPP, June 2024.