

---

# Deconstructing a Decade of Migration Discourse in the European Parliament

---

Giorgi Gogelashvili\* Samia Haque\* Jakob Kleine\*  
Dennis Stroh\* Quirin Unterguggenberger\*

## Abstract

European politics of the past decade has been shaped by the rise of right-wing populism and the increasing politicisation of immigration. Using topic modeling and semantic embeddings, we analyze migration debates in the European Parliament (EP) from 2014 to 2024. Far-right parties disproportionately introduce migration into non-migration debates. Semantic analyses reveal stable rhetorical differences in far-right speech and a stronger alignment with established anti-immigration narratives. These results provide quantitative evidence that the populist right employs a distinct style of migration rhetoric.

## 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). While populist rhetoric in general is characterized by a constructed antagonism between a “pure people” and a “corrupt elite” (Mudde, 2007), right-wing populism in particular is closely tied to the issue of immigration: far-right parties have played a central role in its growing politicisation (Hutter & Kriesi, 2022), which represents a crucial factor of their political success (Kende & Krekó, 2020).

Over the past decade, immigration has become an increasingly salient issue in European election campaigns (Dekeyser & Freedman, 2023) and media coverage (Greussing & Boomgaarden, 2017), especially since the onset of the so-called “refugee crisis” in 2015 (Brücker et al., 2020). In light of these trends, we ask  $Q_1$ : *How prevalent is migration in EP debates and do parties strategically put it on the agenda?*

Electoral gains of populist parties have affected the nature

---

\*Equal contribution . Correspondence to: GG <giorgi.gogelashvili@student.uni-tuebingen.de>.

Project report for the “Data Literacy” course at the University of Tübingen, Winter 2025/26 (Module ML4201). Style template based on the ICML style files 2025. Copyright 2025 by the author(s).

of parliamentary discourse (Schwalbach, 2023). A recent quantitative analysis of EP speech embeddings identified a gradual rise in emotional rhetoric from 1999 to 2022, with right-wing populist groups leading the trend (Subtil & Verger, 2024). In the German national parliament, an LLM-based study found increasing anti-solidarity messaging around immigration, not only for right-wing, but also christian-conservative and liberal parties (Kostikova et al., 2025). Seeing these underlying shifts in rhetoric, we investigate  $Q_2$ : *How does migration rhetoric differ between parties and how do these patterns evolve over time?*

More fine-grained analyses revealed the use of common underlying narratives, defined as “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” (Seiger et al., 2025).  $Q_3$ : *Do parties resort to known anti-immigration narratives?*

Addressing these questions clarifies how the growing prominence of right-wing populism and immigration as a salient political issue manifest in EP debates. Analyzing all parliamentary speeches between 2014 and 2024, we employ unsupervised topic modeling to break down the relative prevalence of migration discourse and provide evidence of agenda-setting strategies by right-wing populist groups (Section 3.1). By embedding migration-related speeches, we examine semantic dimensions that differentiate party groups (Section 3.2) and identify higher use of anti-immigration narratives among right-wing speakers (Section 3.3).

## 2. Data and Methods

### 2.1. Dataset description

We use the ParlLawSpeech dataset (Schwalbach et al., 2025), which contains the full verbatim transcriptions of parliamentary debates between 1999 and 2024 with rich metadata. We restrict the analyses to the last two complete legislative periods (2014 to 2024), as the 2015/2016 “refugee crisis” could mark a qualitative shift in migration debates that might bias topic-model estimation. In a preliminary test, evidence for this effect was reflected by an increase in migration-labeled speeches from  $\approx 1\%$  before 2014 to over 2.5% thereafter.

To contextualize and validate our findings around rhetorical patterns with an external benchmark, we use party positions from the Chapel Hill Expert Survey (CHES; Rovny et al., 2025), which provides repeated cross-sectional ratings of party ideology (e.g. left-right spectrum) and policy positions (e.g. on immigration) using standardized scales. CHES items were rescaled to a common 0–10 metric where necessary, following Adams et al. (2014). National party affiliations were taken from the EP’s Open Data Portal for dataset merging. CHES ratings from 2014, 2019, and 2024 were used for correlations with semantic speech measures.

In 2015 and 2016, the total number of speeches increases sharply (over 50,000), compared to an average of  $10,098 \pm 2,724$  in other years. This phenomenon is largely explained by changes in parliamentary procedure adopted at the end of 2016 (Tremblay-Antoine et al., 2024), most notably the discontinuation of short written statements. For the main analyses, we retain written declarations, as they often contain substantive party positions. Because such declarations could be co-signed by multiple Members and thus appear duplicated in the dataset, we retained only the first instance.

## 2.2. Data preprocessing

We use English translations to control for potential biases introduced by embeddings of different languages. Official translations were provided until 2012, which were extended with machine translations by Parllaw until 2019 (Rauh, 2023). We filled in the remaining translations using *Gemini 2.5-flash* (Gemini Team, 2025) and tested its translations on a random sample of speeches translated by Parllaw. We checked that it (i) preserves speeches which are in English and (ii) creates translations with embeddings close to Parllaw’s (lower 0.95 bootstrap CI bound of mean cosine similarity: 0.969;  $n = 1001$ ).

The speeches include bracketed comments that contain markers of the original language, parliamentary incidents, or procedural notes. We removed these comments using rule-based methods. There are also substantial redundancies at the start and end of speeches (greetings, etc.). These sections follow similar rhetorical structures but exhibit substantial lexical variation. To remove redundant formalities, we identified low-impact sentences using TF-IDF, which scores the amount of information of a sentence (Sparck Jones, 1972). Two raters annotated 100 sampled low-scoring sentences for informational relevance. With these ratings, we fit a logistic regression model to estimate the threshold required for 95% accuracy. For each speech, the first and last sentences were removed if their TF-IDF score was below this threshold.

Due to the fragmentation of right-wing forces in the EP and realignment across legislative terms, parliamentary groups were sorted into six ideological blocs following Kaiser & Mittag (2023), see Figure 1.

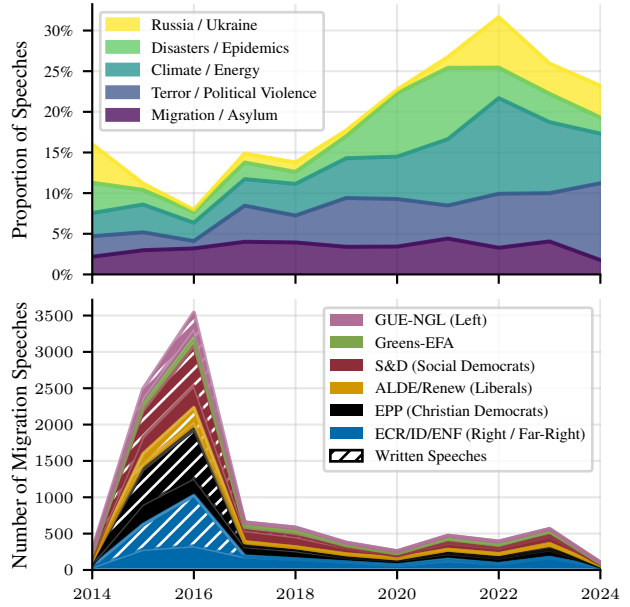


Figure 1. LDA results. **Top:** Number of speeches for selected topics in EP debates, divided by total number of speeches per year. Topic labels were created manually based on most-frequent topic words. See repository for an interactive version with all topics. **Bottom:** Absolute number of migration speeches by parliamentary group. Written speeches were discontinued in 2017.

## 2.3. Methods

### 2.3.1. TOPIC MODELLING WITH LDA

To isolate migration-related discourse from all speeches, we use *Latent Dirichlet Allocation* (Blei et al., 2003). LDA models each speech as a probabilistic mixture of a predefined number of topics, where each topic is defined by a distribution over words. For every speech, the model estimates the probability of belonging to each topic.

We evaluated multiple model specifications and selected the model based on topic coherence (final score: 0.56) and manual inspection of topic interpretability. The selected model comprises 30 topics, one of which assigned the highest probabilities to the words *refugee*, *border*, and *migration*. Speeches were categorized as migration-related if they had an above-threshold probability for this topic ( $n = 9705$ ).

The threshold was determined by two raters sampling 100 speeches from the probability range where the cutoff was expected based on initial tests and classifying whether they were migration-related. Using receiver operating characteristic analysis, we identified the threshold that minimized the difference between true and false positive rates.

### 2.3.2. QUANTIFYING AGENDA SETTING

Seeing that parties strategically influence the debate (Saldivia Gonzatti & Völker, 2026), we investigate whether parties systematically emphasize migration in different contexts and thereby “put it on the agenda”. We look at debates where all migration-related speeches stem from the same party bloc, indicating that migration is fed into an unrelated discussion by that group. We call such a debate an instance of agenda-setting and assign it a ‘true’ topic: the topic with the highest probability among its speeches.

### 2.3.3. SEMANTIC EMBEDDINGS

A pool of candidate embedding models was selected from the MTEB leaderboard (Enevoldsen et al., 2025) based on overall performance and parameter count. Because general-purpose semantic embeddings may primarily capture stylistic and topical variations, we tested that ideological difference is meaningfully encoded in the embeddings. For each candidate model, we compared intra- and interparty similarity distributions using a two-sample Kolmogorov-Smirnov test with Bonferroni correction ( $\alpha^* = 0.05/8$ ). All models show significant distributional differences.

To best capture political orientation in the embeddings, we selected the final model (EmbeddingGemma; Schechter Vera et al. (2025)) based on (i) intra- and interparty cosine similarities, (ii) predictive performance of a logistic regression model with political affiliation as target variable, and (iii) K-means clustering quality (homogeneity and completeness) with respect to party membership.

Dimensionality reduction has been used to study temporal ideological change and to recover latent political dimensions (Rheault & Cochrane, 2020). Exploratory PCA revealed that party affiliations have a minor effect on our embedding variance. To identify a subspace in which political differences become more salient, we use Partial Least Squares (PLS), which extracts directions maximally associated with party labels. To mitigate temporal bias, we excluded written speeches (leaving  $n = 5,433$ ) and evaluate the model using temporally adapted leave-one-out cross-validation (Roberts et al., 2017).

The prevalence of established migration-related rhetoric is assessed using semantic search. We use 30 migration narratives identified in a recent report by the European Commission’s Joint Research Centre (Seiger et al., 2025, p.130), organized into four broader “supernarratives.” Each narrative was represented by a short descriptive sentence and embedded. Semantic proximity between narratives and migration speeches ( $n = 9705$ ) is quantified using cosine similarity. Similarity scores are averaged across narratives within each supernarrative to capture high-level trends. To control for keyword-driven effects, we additionally con-

structed a control narrative at the presumed opposite end of the semantic spectrum (“We need to respect humanitarian principles in handling migration”). To validate that narrative similarity captures meaningful political differences, we correlate similarity scores with the CHES-ratings of party positions (Rovny et al., 2025). Temporal trends and party-bloc differences in narrative prevalence are analysed using linear mixed-effects models with random intercepts and slopes at the party-bloc level. Significance levels were Bonferroni-adjusted to the number of computed Pearson correlations ( $p = 0.05/246$ ) and linear models ( $p = 0.05/6$ ), respectively.

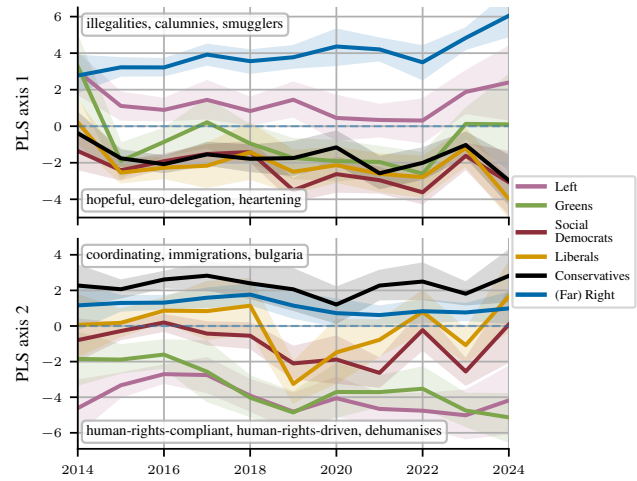


Figure 2. Political bloc embeddings in reduced PLS space over time. Shading denotes bootstrapped 95% confidence intervals of mean yearly embeddings.

## 3. Results

### 3.1. Prevalence of migration speeches & agenda setting

LDA organized the speeches into 30 interpretable and broadly coherent topics, with temporal trends in prevalence roughly consistent with salient world events, such as the outbreak of the Covid-19 pandemic in 2020 or the Russian invasion of Ukraine in 2022 (Figure 1). Although the proportion of migration-related speeches per year remains relatively stable, absolute counts peak in 2015 and 2016, largely due to the use of written declarations (see Section 2).

Migration-related speech rarely occurs in debates where only one party bloc addresses migration. Far right parties are a clear exception, with such instances accounting for  $15.9\% \pm 7.7\%$  of their yearly migration-related speech, compared to  $6.6\% \pm 0.6\%$  for others (ANOVA  $p < .001$ ). They introduce migration mostly in debates where the ‘true’ topic is “Trade”, “EU Security”, or “Economic Development”.

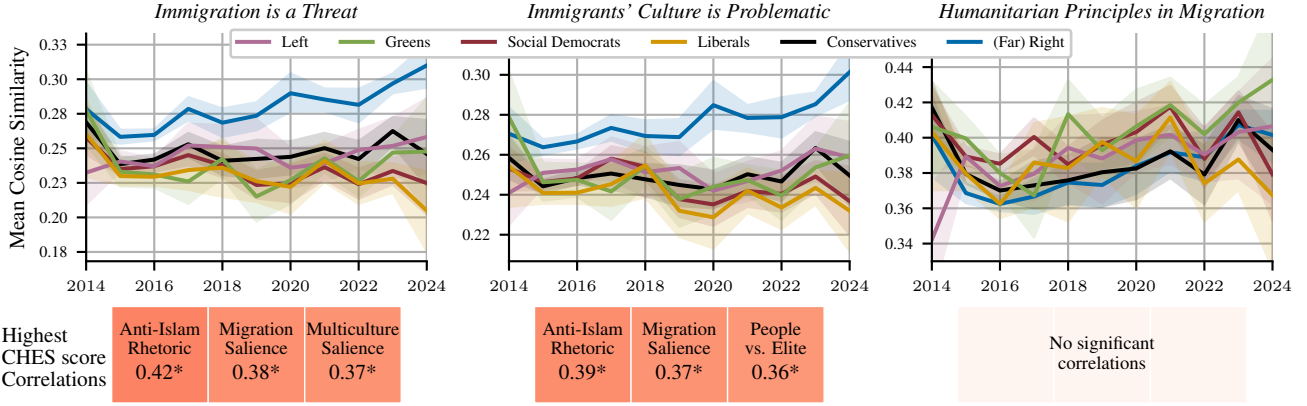


Figure 3. Semantic similarity to two migration supernarratives and the control prompt: temporal trajectories and top correlations to expert party ratings (CHES scores). Shading denotes bootstrapped 95% confidence intervals of mean similarity.

### 3.2. Dimensions separating party blocs

PLS identifies directions in the embedding space that separate the party blocs meaningfully, with an averaged, cross-validated F1 score of  $0.45 \pm 0.35$  (mean  $\pm$  std. across validation folds). The most extreme examples of the corpus vocabulary, as embedded and reduced to two-dimensional space, are displayed in Figure 2 and used for axis interpretation (see Section 4).

### 3.3. Similarities to Established Migration Narratives

Mean semantic similarities between speeches and two exemplary anti-immigration supernarratives, as well as the control prompt, are shown in Figure 3. Mixed linear models reveal party differences for the three anti-immigration supernarratives *immigration is a threat*, *immigrants' culture is problematic*, and *immigration is a burden*, which are significantly higher for far-right speakers (all  $s_{\cos} = 0.27$ ), compared to all other blocs (for *threat* and *burden*  $s_{\cos} \in [0.23, 0.24]$ ; for *problematic culture*  $s_{\cos} \in [0.24, 0.25]$ , all  $p < .003$ ). In contrast, no consistent party differences emerge for narratives framing *immigrants as victims*, the *Us vs. Them* supernarrative, or the control prompt. Across all supernarratives, no significant temporal trends are observed.

Out of all CHES expert party ratings, similarities to the three anti-immigration tropes are most correlated with anti-Islam rhetoric ( $r \in [.39, .45]$ ) and salience of immigration in the party's agenda ( $r \in [.37, .40]$ ), followed by the salience of multiculturalism (for *threat*:  $r = .37$ ; for *burden*:  $r = .39$ ) and populist people vs. elite positioning (for *problematic culture*:  $r = .36$ ). *Us vs. Them* narratives correlate most strongly with their conceptual counterparts, the people vs. elite rating ( $r = .35$ ) and anti-elite salience ( $r = .31$ ).

## 4. Discussion & Conclusion

Our analysis of European Parliament speeches (2014 to 2024) reveal systematic differences in migration rhetoric.

Far-right parties appear to employ agenda-setting behavior by introducing migration-speech into debates, particularly about the economy and security. For other party blocs, this phenomenon is too rare to assert a definite pattern and could partly represent LDA misclassifications.

Dimensionality reduction reveals two axes that separate parties in semantic space. The first axis can be interpreted as *conciliatory*  $\Leftrightarrow$  *oppositional* mode of migration discourse. The second axis is less distinct and is driven by extreme cases referencing security-related concerns, particularly in an Eastern European context, suggesting a *moral*  $\Leftrightarrow$  *pragmatic-benefits / security* dimension. Despite noise in general-purpose embeddings, party blocs separate consistently along both axes, indicating that PLS can recover meaningful latent political dimensions.

Semantic similarity results indicate that far-right rhetoric in the EP is characterized by a higher adoption of previously identified anti-immigrant narratives, without evidence of growth over the last decade. High correlations with established political scales support the validity of this approach as a computational tool for quantifying rhetorical tropes. At the same time, the embedding approach to political speech analysis cannot fully distinguish between mere invocation of statements and actual evaluative stances.

Across all analyses, right-wing populist speech emerges as distinct from other parties. In particular, parties on the political right appear more likely to foreground migration-related themes, while placing less emphasis on humanitarian concerns and framing immigrants in terms of threat, burden, and illegality.



---

## Contribution Statement

Giorgi Gogelashvili worked on data cleaning, analyzing and selecting the embedding model, and dimensionality reduction.

Samia Haque conducted preliminary data analysis, literature review and report editing.

Jakob Kleine implemented speech translation, LDA, analyzed agenda setting, and organized the repository.

Dennis Stroh prepared and merged the Parllaw and CHES datasets, and worked on data cleaning.

Quirin Unterguggenberger co-evaluated and visualized LDA, analyzed narrative similarities, and wrote the majority of the abstract and introduction.

All authors conducted initial literature review and jointly wrote the report.

## Note

All code for this project, an interactive LDA plot, and instructions for data retrieval can be found in the following repository: [https://github.com/derkleinejakob/euparl\\_discourse](https://github.com/derkleinejakob/euparl_discourse).

## References

- Adams, J., Ezrow, L., and Somer-Topcu, Z. Do voters respond to party manifestos or to a wider information environment? An analysis of mass-elite linkages on European integration. *American Journal of Political Science*, 58(4):967–978, 2014.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent Dirichlet Allocation. *Journal of machine Learning research*, 3(Jan): 993–1022, 2003.
- 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.
- Dekeyser, E. and Freedman, M. Elections, Party Rhetoric, and Public Attitudes Toward Immigration in Europe. *Political Behavior*, 45(1):197–209, 2023.
- Enevoldsen, K., Chung, I., Kerboua, I., Kardos, M., Mathur, A., Stap, D., Gala, J., Siblini, W., Krzemiński, D., Winata, G. I., Sturua, S., Utpala, S., Ciancone, M., Schaeffer, M., Sequeira, G., Misra, D., Dhakal, S., Rystrom, J., Solomatin, R., Ömer Çağatan, Kundu, A., Bernstorff, M., Xiao, S., Sukhlecha, A., Pahwa, B., Poświata, R., GV, K. K., Ashraf, S., Auras, D., Plüster, B., Harries, J. P., Magne, L., Mohr, I., Hendriksen, M., Zhu, D., Gisserot-Boukhlef, H., Aarsen, T., Kostkan, J., Wojtasik, K., Lee, T., Šuppa, M., Zhang, C., Rocca, R., Hamdy, M., Michail, A., Yang, J., Faysse, M., Vatolin, A., Thakur, N., Dey, M., Vasani, D., Chitale, P., Tedeschi, S., Tai, N., Snegirev, A., Günther, M., Xia, M., Shi, W., Lù, X. H., Clive, J., Krishnakumar, G., Maksimova, A., Wehrli, S., Tikhonova, M., Panchal, H., Abramov, A., Ostendorff, M., Liu, Z., Clematide, S., Miranda, L. J., Fenogenova, A., Song, G., Safi, R. B., Li, W.-D., Borghini, A., Cassano, F., Su, H., Lin, J., Yen, H., Hansen, L., Hooker, S., Xiao, C., Adlakha, V., Weller, O., Reddy, S., and Muennighoff, N. Mmteb: Massive multilingual text embedding benchmark. *arXiv:2502.13595*, 2025.
- 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, 2023.
- Gemini Team. Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities, 2025. *arXiv:2507.06261*.
- 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, 2017.
- Hutter, S. and Kriesi, H. Politicising immigration in times of crisis. *Journal of Ethnic and Migration Studies*, 48(2): 341–365, 2022.
- Kaiser, W. and Mittag, J. Seventy years of transnational political groups in the European Parliament. *European Parliamentary Research Service*, 2023.
- Kende, A. and Krekó, P. Xenophobia, prejudice, and right-wing populism in East-Central Europe. *Current Opinion in Behavioral Sciences*, 34:29–33, 2020.
- 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, 2025. *arXiv:2509.07274*.
- Mudde, C. *Populist Radical Right Parties in Europe*. Cambridge University Press, Cambridge, 2007.
- Rauh, C. ParlLawSpeech documentation — machine translation of parliamentary speeches, 2023. URL <https://html-preview.github.io/?url=https://github.com/ChRauh/OPTED-WP5-Translation/blob/main/PLS-TranslationDocumentation.html>.
- Rheault, L. and Cochrane, C. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Polit. Anal.*, 28(1):112–133, 2020.

- 
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.
- Rovny, J., Polk, J., Bakker, R., Hooghe, L., Jolly, S., Marks, G., Steenbergen, M., and Vachudova, M. A. The 2024 Chapel Hill Expert Survey on political party positioning in Europe: Twenty-five years of party positional data. *Electoral Studies*, 97:102981, 2025.
- 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, 2017.
- Saldivia Gonzatti, D. and Völker, T. Far-right agenda setting: How the far right influences the political mainstream. *European Journal of Political Research*, 65(1):101–123, 2026.
- Schechter Vera, H., Dua, S., and Team, E. Embeddinggemma: Powerful and lightweight text representations. *arXiv:2509.20354*, 2025.
- 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 Seböck, M. ParlLawSpeech. (Version 1.0.0) [Data set]. GESIS, Cologne., 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.
- Sparck Jones, K. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.
- Subtil, H. and Verger, V. Emotional rhetoric and the rise of populism in the European parliament. Policy Brief 108, IPP, 2024.
- Tremblay-Antoine, C., Jacob, S., Dufresne, Y., Poncet, P., and Dinan, S. An open window into politics: A structured database of plenary sessions of the European parliament. *European Union Politics*, 25(3):605–622, 2024.