Deconstructing a Decade of Migration Discourse in the European Parliament

Giorgi Gogelashvili [*]   **Samia Haque** [*]   **Jakob Kleine** [*]
Dennis Stroh [*]   **Quirin Unterguggenberger** [*]

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).

This increasing politicisation can be understood through *agenda-setting theory*, which examines how political actors influence which issues receive attention (McCombs & Shaw, 1972). By strategically emphasizing migration in parliamentary debates, populist parties actively shape the political agenda, potentially pulling centrist parties into their discursive framework.

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 from 1999 to 2022, 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).

These previous successes in deconstructing and quantifying highly impactful political trends inspired the present investigation of how the growing prominence of right-wing populism and immigration as a salient political issue manifest in EP debates, with potential implications for broader societal discourse and legislative outcomes. All parliamentary speeches between 2014 and 2024 as recorded by the ParLawSpeech dataset inform the analyses.

We employ unsupervised topic modeling to break down the prevalence of migration debates relative to other topics

[*]Equal contribution . Correspondence to: AB <first1.last1@student.uni-tuebingen.de>.

over the last decade, and provide evidence of agenda-setting strategies employed by right-wing populist groups (Section 2). By embedding migration-labeled speeches into vector representations, we examine the semantic dimensions along which party groups can be differentiated and identify [wording too strong?] an increased use of previously identified anti-immigration narratives by right-wing groups [compared to moderate factions?] (Section 3).

2. Data and Methods

2.1. Dataset description

We are using the novel *ParlLawSpeech* (PLS) dataset from Schwalbach et al. 2025 for the investigation of our study. It contains more than 570,000 plenary speeches from legislative periods of the European parliament (EP) between 1999 and 2024. The authors also provide (partially) machine translated text in English for about 40% of the speeches, since the EP stopped providing official translations around the end of 2012. Furthermore, the dataset contains metadata on the speakers and the speeches given, e.g. date and agenda item under which the speech was given, if submission was in written form and/or from multiple *members of parliament* (MEPs), or the speaker's party affiliation (referring to European political parties/groups), among other. We further enriched the dataset with metadata accessible from the public API of the EP's "Open Data Portal", in particular the national party affiliations of each speaker (by using the *EP-ID* of the respective MEPs). This allowed us to link the PLS dataset with the *Chapel Hill Expert Survey* (CHES) from Rovny, Bakker et al. 2025. The CHES dataset estimates party positioning on European integration, ideology (e.g. left/right) and policy issues for national parties in all member states of the European Union (EU). The study surveyed hundreds of experts roughly every four years between 1999 and 2024 and more recently (**TODO**: since when???) also includes ratings of non-EU policy issues such as immigration or anti-elite rhetoric (**TODO:** which are relevant in particular?) Assuming that the ideological orientation of a speaker's affiliated national party roughly reflects his own position, the CHES data set could help us to better control our analyses, as membership of a European party (group) presumably allows for less detailed/granular statements/assumptions. We restrict our analyses to the last two complete legislative periods (2014–2024), as the 2015/2016 "refugee crisis" marks a qualitative shift in the nature and salience of migration-related debates that could bias topic-model estimation. This shift is reflected in the proportion of migration-labeled speeches, which remains around 1% prior to 2014 but rises to over 2.5% thereafter.

2015 and 2016 anomaly In 2015 and 2016, there is a drastic increase in the number of speeches compared to the other years (72,964 in 2015 and 16 on average, compared to 10,098 in the other years on average). This difference can be explained with a rule-change that was adopted by the parliament at the end of 2016, discontinuing so-called 'written declarations' that allowed party members to hand in short expressions of opinion on a certain issue [TODO ref]. Omitting all 'written' speeches (TODO: explain what written flag means) can normalize the number of speeches per year. However, for the general analysis, we keep written declarations as part of the dataset because they can contain relevant stances of the parties on political issues. Since written declarations could be co-signed by multiple speakers, the can appear duplicated in the dataset. These duplicated items were removed, keeping only the first instance of duplicated a speech.

2.2. Data preprocessing

Handling Duplicates (TODO)

Removing Commentary (TODO) We detect high amount of superfluous commentary in transliterated speeches: markers of the original language, background incidents, and procedural notes. These markers might be source of unwanted bias, which we want to avoid. Fortunately they are predominantly located within parentheses and can be easily removed with rule-based methods. We also observe substantial redundancy in the openning and closing sections of the speeches. These sections follow similar rhetorical structures but exhibit substantial lexical variation. To identify low-impact sentences we use TF-IDF algorithm to score the amount of information they contain. We construct separate corpora for opening and closing sentences, and an average TF-IDF score is computed for each sentence. Two independent raters annotated a sample of 100 low-scoring sentences for informational relevance. We then fit a logistic regression linking TF–IDF score percentiles to these annotations and used the model to estimate the threshold required to achieve 95% classification accuracy.

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)*

---

[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 En-

created translations whose embeddings are very similar to Parllaw's (bootstrapped 0.95 confidence interval of mean 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, we use Latent Dirichlet Allocation *(LDA)* [TODO: source]. LDA fits a probabilistic model on a corpus of documents (here: our parliamentary speeches) and identifies a fixed number of topics in the corpus which are characterized by words that have a high probability of appearing in the topic. For each speech and topic, it assigns a probability of that speech belonging to the topic.

We compared different model fits and chose one based on the its coherence score (final coherence: XX) and throguh manually inspecting the fidelity of the created topics. Our final model contains 30 topics, one of which assigned highest probabilities for the words X, Y, Z (TODO words), which we call "migration topic".

To determine an appropriate migration-topic probability threshold, two raters sampled 100 speeches from the probability range where the cutoff was expected based on initial tests (prob = [0.20, 0.35]) and classified whether they were migration-related. We then used receiver operating characteristic (ROC) analysis to identify the threshold that minimized the difference between true and false positive rates (prob = 0.25).

(TODO finish this)

2.3.2. SEMANTIC EMBEDDINGS

Semantic embeddings have been widely used in political text analysis (Miok et al., 2024; Nanni et al., 2021; Rudkowsky et al., 2018). Our aim is to capture patterns in how different political groups address migration. We select candidate embedding models from the MTEB leaderboard (Enevoldsen et al., 2025), based on overall performance and parameter count. Final model selection is based on (i) intra- and interparty cosine similarities, (ii) predictive performance of

---

a logistic regression model with political affiliation as our target variable, and (iii) Kmeans clustering quality measured by homogenity and completeness. Based on these metrics we have selected google/embeddinggemma-300m (Schechter Vera et al., 2025) as our final embedding model, all analysis using semantic embeddings are conducted with this model.

A key concern is that general-purpose semantic embeddings may be primarily capturing stylistic and topical variations and subsequently political group ideologies influence on the embeddings might be neglegible. We test whether intra- and interparty similarity distributions differ substantially with a two-sample Kolmogorov-Smirnov test for each candidate model. We apply Bonferroni correction accross 8 models ($\alpha = 0.05$, $\alpha^* = 0.05/m$). All models showed significant distributional differences, with test statistic $\mathcal{D}$ ranging between 0.058 and 0.1.

We examine whether party affiliations are encoded in speech embeddings and how these patterns evolve over time. Dimensionality reduction has been used to ascertain parties ideological shift over time and to reveal underlying political dimension with word associations for each reduced axis (Rheault & Cochrane, 2020). Exploratory analysis showed that, although party influence is present, it is not the defining factor of our semantic embeddings. To better understand how party affiliations manifest in the vector space, we aim to identify a subspace of the embedding space in which political and ideological differences become more salient.

To this end, Instead of simply using PCA, we employ Partial Least Squares (PLS). PLS allows us to find directions in the embedding space that are maximally associated with party labels, making it suitable for uncovering latent political dimensions that are not necessarily dominant in the overall variance of the data.

The prevalence of established migration-related rhetoric was assessed using semantic search in a shared embedding space. We used 30 suitable migration narratives that were identified in a recent report by the European Commission's Joint Research Centre (Seiger et al., 2025, p.130), which were organized into four broader categories or "supernarratives" [list them here?]. Each narrative was represented by a short descriptive sentence, which was embedded using the model's built-in "retrieval-query" prompt.

Semantic proximity between narratives and speeches was quantified using cosine similarity. Semantic similarity scores were averaged across all narratives belonging to a supernarrative to identify high-level trends. We additionally constructed one comparative narrative at a presumably opposed end of the rhetoric spectrum ("We need to respect humanitarian principles in handling migration") to ensure that potential party differences cannot be caused by the

frequency of certain keywords like "migration" alone.

To validate whether semantic similarity to these narratives captured meaningful political differences, we correlated similarity scores with expert-coded party positions on migration policy and overall ideology from the Chapel Hill Expert Survey (Rovny et al., 2025). Pearson correlation coefficients were evaluated using a Bonferroni-adjusted significance threshold to account for multiple comparisons. Temporal trends and party-block differences in narrative prevalence were analysed as fixed effects of linear mixed-effects models, which incorporated random intercepts and slopes at the party-block level. Because a separate model was fit for each of the six supernarratives, we Bonferroni-corrected significance levels to $p = 0.05/6$.

3. Results

3.1. Salience of migration-related speech

3.1.1. WHO TALKS ABOUT MIGRATION?

Figure [TODO] shows the prevalence of migration as parliamentary topic, compared to selected reference topics. It shows a consistent prevalence of the issue since 2014 with minor fluctuations, indicating a continuous interest in the topic.

3.1.2. SETTING THE AGENDA

A phenomenon studied in the literature is *agenda setting*, which describes a political acteur's strive to place their topic of interest in the discourse [TODO: ref]. In the context of parliamentary debate in the EU, it is interesting to ask: do parties try to put migration on the agenda? For this purpose, we looked at debates where all speeches that were labelled as migration-related stem from the same party block, indicating that migration is not the main topic of the debate, but continuously mentioned by that party block. We defined such a debate as an instance of agenda-setting by the respective party. All parties exhibit agenda-setting behaviour. However, it is drastically more pronounced for far-right parties, with X% of their overall migration-related speech occuring debates where they are the only party talking about migration — compared to Y% for other parties on average [TODO: add significance statement]. Do parties systematically bring up migration in certain contexts? To answer this, we assigned each debate a 'true' topic, which is the topic with highest average probability among all speeches in that debate. Far-right parties bring up migration mostly in debates where the 'true' topic is related to foreign relations ('Trade Relations', 'EU Security / Defense', or 'Economic Development') while progressive parties (social democratic, green, left) link it most to humanitarian situations ('Disasters / Epidemics', 'International Conflicts', and 'Debate Etiquette / Brexit'), and for liberal and conservative there is

no clear pattern ('EU Security / Defense', 'Debate Etiquette / Brexit', and 'Economic Development').

[TODO: add example?] [TODO: add interpretation?]

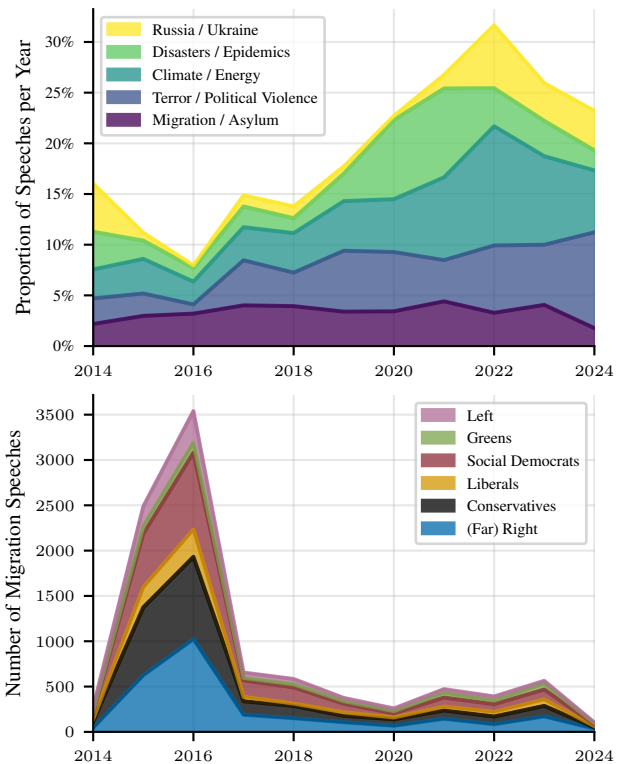*Figure 1.* **Top:** Prevalence of selected topics in European Parliament debates over the past decade, as identified by LDA topic modeling. Proportions are computed by dividing by the total number of speeches per year. See repository for an interactive version with all topics. **Bottom:** Absolute number of migration speeches by parliamentary group.

3.2. Interpreting semantic differences

To mitigate and quantify temporal bias in our PLS analysis, we excluded written speeches predominantly occurring in 2015 and 2016 and fitted the PLS model on the remaining 5,433 observations. Furthermore, we employed a leave-one-out cross-validation strategy adapted for temporal data (Roberts et al., 2017) The resulting micro-averaged F1 score was 0.45 ± 0.35 (mean $\pm$ std. accross validation folds). While a clear interpretation of the underlying political dimensions requires substantial domain knowledge, we believe that combining word associations with extreme examples of speeches along each cardinal direction provides some clues about their connotations. Based on this analysis, we interpret the first PLS axis as a **conciliatory** $\Leftrightarrow$ **oppositional** discourse spectrum, although it's harder to discern underlying political dimension for the second axis, we pre-
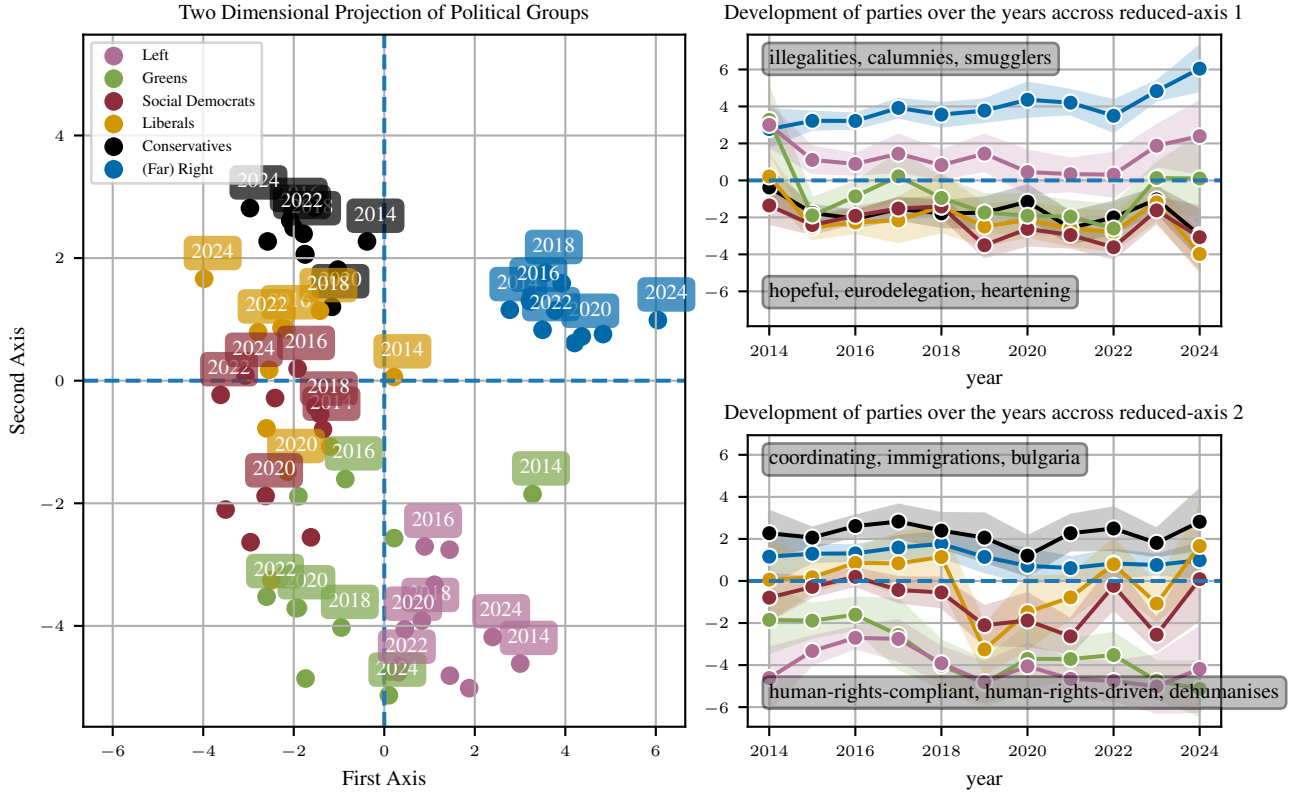
Figure 2. **Left:** Position of each political group. **Right:** Movement of political groups over the time displayed separately for each dimension.

sume it to be **moral / human-rights ⇔ pragmatic-benefits** debate Figure 2. [TODO CHESS correlations]

3.3. Similarities to Established Migration Narratives

Mean semantic similarities of speeches to two exemplary anti-immigration supernarratives and our constructed humanitarian prompt are visualized in Figure 3. Mixed linear models revealed party differences for the three anti-immigration supernarratives "immigration is a threat", "immigrants' culture is problematic', and "immigration is a burden", which were significantly higher for far-right speakers (all $d_{cos} = 0.27$), compared to all other blocks (for threat & problematic $d_{cos} = 0.25$, for burden $d_{cos} = 0.24$, all $p < .003$). For narratives falling under "immigrants as victims" or the populist "Us vs. Them", and the comparison "Humanitarian" prompt, no consistent differences emerged. For all supernarratives, no significant temporal trends were found.

Out of all expert party ratings, similarities to the three anti-immigration tropes were most correlated with anti-Islam rhetoric ($r = [.39, .45]$) and salience of immigration in their political agenda ($r = [.37, .40]$), on third position followed salience of multiculturalism ($r_{threat} = .37, r_{burden} = .39$)

and populist people vs. elite ($r_{problem} = .36$). Immigrants as victims framing had lower correlations with similar dimensions, the us vs. them narrative category correlated most with its equivalent people vs. elite rating ($r = .35$). The humanitarian comparator yielded no significant CHES score correlations.

4. Discussion & Conclusion

Our analysis of European Parliament speeches (2014–2024) reveals systematic patterns in migration discourse. Topic modeling shows right-wing parties disproportionately drive migration salience while strategically introducing it into unrelated debates—particularly economic and security discussions—demonstrating active agenda-setting behavior. Partial Least Squares analysis indicates semantic convergence between centrist and right-wing rhetoric following the 2015 refugee crisis, suggesting diffusion of populist framing.

Our 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 indicate that this approach is a viable and promising computational tool for quantifying rhetoric tropes.
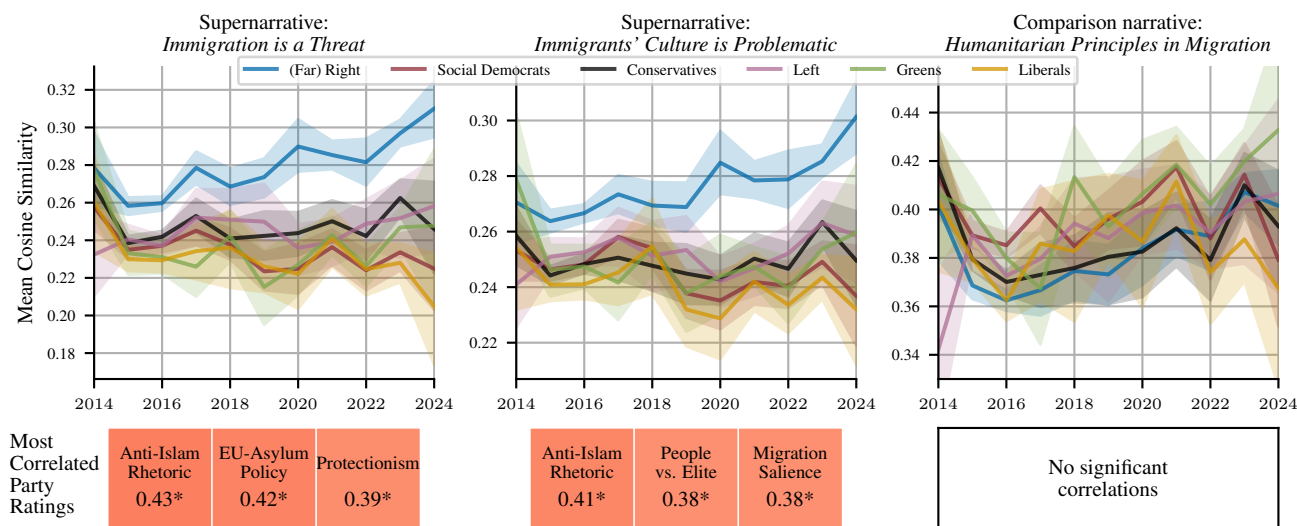
Figure 3.

Methodologically, we demonstrate how combining LDA topic modeling, semantic embeddings, and narrative detection can capture both topical focus and rhetorical framing. While translation consistency and temporal confounding present limitations, our findings illuminate how populist discourse enters mainstream parliamentary institutions, with implications for European democratic deliberation.

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.

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., Rystrøm, 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 preprint arXiv:2502.13595*, 2025. doi: 10.48550/arXiv.2502.13595. URL https://arxiv.org/abs/2502.13595.

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].

McCombs, M. E. and Shaw, D. L. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.

Miok, K., Hidalgo Tenorio, E., Osenova, P., Benítez-Castro, M.-A., and Robnik-Sikonja, M. Multi-aspect multilingual and cross-lingual parliamentary speech analysis. *Intelligent Data Analysis*, 28(1):239–260, February 2024. ISSN 1571-4128. doi: 10.3233/ida-227347. URL http://dx.doi.org/10.3233/IDA-227347.

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.

Nanni, F., Glavas, G., Rehbein, I., Ponzetto, S. P., and Stuckenschmidt, H. Political text scaling meets computational semantics. *ACM/IMS Transactions on Data Science*, 2(4):1–27, November 2021. ISSN 2691-1922. doi: 10.1145/3485666. URL http://dx.doi.org/10.1145/3485666.

Rheault, L. and Cochrane, C. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Polit. Anal.*, 28(1):112–133, January 2020.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-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. doi: https://doi.org/10.1111/ecog.02881. URL https://nsojournals.onlinelibrary.wiley.com/doi/abs/10.1111/ecog.02881.

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.

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, S., and Sedlmair, M. More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, 12(2–3):140–157, April 2018. ISSN 1931-2466. doi: 10.1080/19312458.2018.1455817. URL http://dx.doi.org/10.1080/19312458.2018.1455817.

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.

Schechter Vera, H., Dua, S., Zhang, B., Salz, D., Mullins, R., Raghuram Panyam, S., Smoot, S., Naim, I., Zou, J., Chen, F., Cer, D., Lisak, A., Choi, M., Gonzalez, L., Sanseviero, O., Cameron, G., Ballantyne, I., Black, K., Chen, K., Wang, W., Li, Z., Martins, G., Lee, J., Sherwood, M., Ji, J., Wu, R., Zheng, J., Singh, J., Sharma, A., Sreepat, D., Jain, A., Elarabawy, A., Co, A., Doumanoglou, A., Samari, B., Hora, B., Potetz, B., Kim, D., Alfonseca, E., Moiseev, F., Han, F., Palma Gomez, F., Hernández Ábrego, G., Zhang, H., Hui, H., Han, J., Gill, K., Chen, K., Chen, K., Shanbhogue, M., Boratko, M., Suganthan, P., Duddu, S. M. K., Mariserla, S., Ariafar, S., Zhang, S., Zhang, S., Baumgartner, S., Goenka, S., Qiu, S., Dabral, T., Walker, T., Rao, V., Khawaja, W., Zhou, W., Ren, X., Xia, Y., Chen, Y., Chen, Y.-T., Dong, Z., Ding, Z., Visin, F., Liu, G., Zhang, J., Kenealy, K., Casbon, M., Kumar, R., Mesnard, T., Gleicher, Z., Brick, C., Lacombe, O., Roberts, A., Sung, Y., Hoffmann, R., Warkentin, T., Joulin, A., Duerig, T., and Seyedhosseini, M. Embeddinggemma: Powerful and lightweight text representations. 2025. URL https://arxiv.org/abs/2509.20354.

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.