

# Divergent Discourse: a Word Embedding Approach to Exploring Issue Framing and Semantics Polarization within the U.S. Congress

## Project Proposal

Sean Hambali

March 20, 2024

## 1 Project overview

Mass polarization studies suggest that voters tend to overstate the degree of polarization that is occurring in the society, a phenomenon known as “false” polarization (e.g., Van Boven, Judd, and Sherman (2012), Levendusky and Malhotra (2016), Enders and Armaly (2019)). Such accentuated perceived polarization is, in part, exogenously influenced by politicians’ use of negative frames (Banks et al. 2021) in policy discourses. Thus, studying the prevalence of issue framing and measuring the extent of upstream partisan polarization among U.S. politicians is crucial to understanding downstream effects on mass polarization. While existing studies on polarization within the U.S. Congress do share a broad consensus of intensifying partisanship in both the Senate and House of Representatives (Barber and McCarty 2015), most of these studies have solely relied on the use of DW-NOMINATE score (Poole and Rosenthal 1985) to measure party- and senator-level ideology score. Using a case study of the 1920s, Caughey and Schickler (2016) demonstrates that this score does not necessarily provide a consistent ideological interpretation over time. Further, few studies have differentiated senators’ partisanship along policy issues, which is an important factor in explaining heterogeneity in policy enactment success across different sectors.

In this research, I plan to complement existing studies by using a word embedding approach to explore partisanship as reflected in U.S. congressional speeches. Similar to Gentzkow, Shapiro, and Taddy (2019), Rheault and Cochrane (2020), Rodriguez, Spirling, and Stewart (2023), I narrow my focus on the *semantic* aspect of partisanship. Word uses can potentially capture unconscious bias and reveal ideological positions in ways that aren’t fully reflected by roll-call votes. Using parsed congressional speech data from the 107th-114th Congress, I aim to describe both the overall and issue-specific trends in semantic polarization. For the latter, I focus on topics that have high degree of mass polarization (Gallup 2017), i.e., immigration, abortion, gun laws, climate change, healthcare, and racial justice. To complement these findings, I also plan to measure senator-level ideology positions on these policy topics, which will provide a nuanced understanding of within-party divergence in policy stances, and to inform whether policy compromises are not out of reach in certain policy issues.

Specifically, my research aims are as follows:

1. Provide quantitative assessments on the overall and sector-specific semantic polarization trend

2. In each policy topic, provide indicative assessments for each political party's issue framing strategies by identifying words that are closest to the topic keywords of interest in the embedding space.
3. In each policy topic, provide quantitative assessments on senator-level positions along the ideological vector space.

## 2 Proposed methodology

### 2.1 Data

For this research, I will be using Gentzkow, Shapiro, and Taddy (2019)'s data on U.S. congressional speeches. Although their original data spans from 1873 to 2016, I will only consider data from the 1981-2017 period (97th-114th Congress), for which speech is available at a daily frequency. The data is publicly accessible and can be downloaded at this [link](#).

### 2.2 Method

To achieve the above-mentioned first research aim, I plan to leverage the use of **GloVe** as a pre-trained embedding model to initialize the embedding values during the training process. Following the approach outlined by Rheault and Cochrane (2020), I fit the model with party-congress indicator variables included as additional metadata to get the embedding value for each party-congress combinations. These embedding values are then fit to a PCA model, and the measure of interest that I will extract from this exercise is the first component scores, which reflects a particular speech's ideological component Rheault and Cochrane (2020). I repeat the same procedure to get the sector-specific trends, but this time fitting the model only on corpus that contains the keywords related to each policy topic. The same implementation will be repeated for achieving the third research aim; I will use topic-specific corpus and include speaker metadata in the model to get their respective embeddings.

The second research goal aims to look at the strategies that parties use to frame political issues. For this, I use the same pre-trained embedding model on party-topic corpus to identify words that are closest to the topic keyword for each party in the embedding space. To determine whether these semantics uses are significantly different along the party lines, I plan to use Rodriguez, Spirling, and Stewart (2023)'s regression-based **conText** package which allows me to draw inferences on statistical significance of the party coefficient. I include an interaction between party and congress session in the regression to determine whether the partisan gap has widened over time.

### 2.3 Parameter of project success

The project success will primarily rely on the successful implementation of the analysis codes. I expect that some of the tasks will be computationally expensive, especially because it involves fitting multiple embedding models to a corpus with millions of words. When the task becomes too computationally prohibitive, I plan to narrow down the study period. I hope to be able to show clear trends of semantics divergence along party lines, which would support the hypothesis of growing partisanship within the U.S. Congress. I also expect that the results would show clear party differences in words associated with each topic keyword, implying that parties tend to frame issues differently, and possibly, increasingly so. Lastly, it would be interesting if the results show clear patterns of within-party divergences on certain policy topics, which would suggest that DW-NOMINATE scores might mask important ideology differences across policy issues.

## References

- Banks, Antoine, Ernesto Calvo, David Karol, and Shibley Telhami. 2021. “#PolarizedFeeds: Three Experiments on Polarization, Framing, and Social Media.” *The International Journal of Press/Politics* 26 (3): 609–34. <https://doi.org/10.1177/1940161220940964>.
- Barber, Michael J., and Nolan McCarty. 2015. “Causes and Consequences of Polarization.” In, edited by Nathaniel Persily, 1st ed., 15–58. Cambridge University Press. <https://doi.org/10.1017/CBO9781316091906.002>.
- Caughey, Devin, and Eric Schickler. 2016. “Substance and Change in Congressional Ideology: NOMINATE and Its Alternatives.” *Studies in American Political Development* 30 (2): 128–46. <https://doi.org/10.1017/S0898588X16000092>.
- Enders, Adam M., and Miles T. Armaly. 2019. “The Differential Effects of Actual and Perceived Polarization.” *Political Behavior* 41 (3): 815–39. <https://doi.org/10.1007/s11109-018-9476-2>.
- Gallup. 2017. “Partisan Differences Growing on a Number of Issues.” <https://news.gallup.com/opinion/polling-matters/215210/partisan-differences-growing-number-issues.aspx>.
- Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. “Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech.” *Econometrica* 87 (4): 1307–40. <https://doi.org/10.3982/ECTA16566>.
- Levendusky, Matthew S., and Neil Malhotra. 2016. “(Mis)perceptions of Partisan Polarization in the American Public.” *Public Opinion Quarterly* 80 (S1): 378–91. <https://doi.org/10.1093/poq/nfv045>.
- Poole, Keith T., and Howard Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29 (2): 357–84. <https://doi.org/10.2307/2111172>.
- Rheault, Ludovic, and Christopher Cochrane. 2020. “Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora.” *Political Analysis* 28 (1): 112–33. <https://doi.org/10.1017/pan.2019.26>.
- Rodriguez, Pedro L., Arthur Spirling, and Brandon M. Stewart. 2023. “Embedding Regression: Models for Context-Specific Description and Inference.” *American Political Science Review* 117 (4): 1255–74. <https://doi.org/10.1017/S0003055422001228>.
- Van Boven, Leaf, Charles M. Judd, and David K. Sherman. 2012. “Political Polarization Projection: Social Projection of Partisan Attitude Extremity and Attitudinal Processes.” *Journal of Personality and Social Psychology* 103 (1): 84–100. <https://doi.org/10.1037/a0028145>.