Final Report: A Replication of *Embedding Regression: Models for Context-Specific Description and Inference* (Rodriguez et al., 2023)

PPOL 6801: Text as Data: Computational Linguistics

Bridgette Sullivan & Maria Bartlett

April 7, 2025

I. Introduction

In *Embedding Regression: Models for Context-Specific Description and Inference*, Rodriguez et al., 2023a highlight the power of word embedding techniques in the capture of textual context. The authors identify two shortcomings in current word embeddings literature. First, because researchers are often interested in highly specific focal words, the corresponding corpus is often relatively small; accordingly, there is a trade-off between the gains of capturing corpus-specific context by a researcher training their own embeddings and the computational intensity of this training process. Second, current embedding techniques generally do not have a mechanism for inference (Rodriguez et al., 2023a). To that end, Rodriguez et al., 2023a propose a framework which makes two innovations:

1. Leverages á la carte embeddings to efficiently produce context-specific embeddings.

2. Utilizes a multilinear regression model specification to regress the context-specific embeddings on covariate(s) of interest in order to conduct inference.

The á la carte method requires using pre-trained embeddings (in this case, GloVe) and a transformation matrix to compute the context-specific embeddings (Rodriguez et al., 2023a). The benefit of this method is computational efficiency and corpus specificity: a corpus-specific transformation matrix is calculated once (or, as in the case of Rodriguez et al., 2023a, an external matrix can be used if deemed appropriate) and is multiplied with the pretrained embeddings (Rodriguez et al., 2023a).

The authors demonstrate how to use and interpret the framework by applying it to a question for which they have a strong prior: Was the word 'Trump' used differently post-2016 election relative to the word 'Clinton' (Rodriguez et al., 2023a)? First, the authors demonstrate the calculation of the corpus-specific embeddings of the focal words 'Trump' and 'Clinton' using the proposed á la carte methodology. Next, they regress these embeddings on covariates of interest using the following specification:

$Focal\_word\_embedding = \beta_0 + \beta_1 Trump + \beta_2 Post\_election + \beta_3 Trump X post\_election + \epsilon$

Because the outcome for each observation is a 300-dimension word embedding, the corresponding betas returned by the regression are also multi-dimensional. For interpretability, the authors propose taking the norm of each matrix in order to present a single coefficient for each beta (Rodriguez et al., 2023a). The authors underscore that the units of the normed betas do not have inherent meaning; however, the magnitude of the corresponding betas in the context of a binary covariate conveys a sense of the degree of difference in focal word embeddings between the two values of the covariate (Rodriguez et al., 2023a). As expected based on their prior, the interaction term between Trump and post-election is positive and statistically significant

at $p < 0.001$ (Rodriguez et al., 2023a). Although we do not detail it further here, it is important to note that the authors' framework can also be used with continuous covariates.

In the remainder of the paper, the authors demonstrate three additional potential use cases for their framework:

1. Do Democrats and Republicans use "politically charged" words to mean different things? How does this compare to differences across gender?

2. How did the meaning of the word "empire" change in the U.S. vs. U.K. post-World War II?

3. Does sentiment change between cabinet members and backbenchers of differing parties for various policy issues?

The authors find that there are distinct differences in ways that each of these groups uses specific words. For example, the authors find that there is a more stark difference in how Democrats and Republicans use the word 'immigration' compared to the difference in the use of the word between men and women (regardless of party) (Rodriguez et al., 2023a). The second case reveals that the U.S. and U.K. used the word 'empire' in about the same way, discussing historical empires, up until 1949 where there is a "structural break" that indicates how the U.S. starts to use "empire" in the context of the Soviet Union to talk about a current "empire" while the U.K. continues to talk about historical empires (Rodriguez et al., 2023a).

II. Similarities & Differences

Case 1: Partisanship, Ideology, and Gender Differences

As in the preceding framework demonstration, the authors utilize the á la carte embedding methodology to train embeddings for a focal word. In this research question, they choose a set of "politically charged" words (as well as a set of stopwords for comparison purposes) to independently use as the focal word of interest. The corresponding regression they run - for example, in the case of the word "immigration" - is:

$Immigration\_embedding = \beta_0 + \beta_1 Republican + \beta_2 Male + \epsilon$

Additionally, the authors run "nearest neighbors" and "nearest contexts" analyses for the focal words. We executed each of these analyses (with minimal updates to the code) and replicated all findings.

Case 2: The Meaning of "Empire"

The main results for this case were the beta values over time for the covariate "Congressional Record" and the nearest neighbors for "empire" in the U.S. and U.K. The regression used to evaluate the differences in covariates in their use of "empire" is:

$Empire\_embedding = \beta_0 + \beta_1 Congressional\_Record + \epsilon$

The beta values over time were consistent with those in the author's results. We see the discussed "structural break" occur right at 1949-1950 that shows the beta values on the covariate to steeply increase, showing that the two groups start to talk about "empires" in different ways. Next, we find the nearest neighbors for "empire" across each time period, finding that these words well represent the author's conclusion. Prior to 1950, words referencing past empires were prevalent for both the U.S. and the U.K., but after 1950 the U.S. started to use words to mainly reference the Soviet Union and the U.K. continues to use words to describe past empires. While the graphic present in the author's paper is helpful to understand specific terms used by each group, it is slightly misleading because these cosine similarity values are on different scales. We decided to change the orientation of our visualization of nearest neighbors from horizontal to vertical to allow for a better understanding of just how dissimilar the post 1950 nearest neighbors are.

III. Autopsy

Outdated Code

Several of the functions used were slightly outdated and required updated arguments to properly run code. One was a very simple fix in `geom_vline()` and `geom_hline()` which required changing the `size` argument to `linewidth` as the argument name had changed. The second appeared to be caused by an update the authors had made in their `conText` package since the replication scripts were last updated: in their replication code, when they use the `conText()` function they set `boostrap` = TRUE for standard error calculation (Rodriguez et al., 2023b). However, the function also includes a `jackknife` option for standard errors that is by default = TRUE (Rodriguez, n.d.). Because these cannot both be set = TRUE, the original code throws an error unless the user adds the line `jackknife` = FALSE (Rodriguez, n.d.). Thus, we added this line to the corresponding uses of `conText()` in order to execute the code properly.

Repetition via Organization

The scripts were well-organized and digestible. While this organization was generally helpful, it also added several challenges: first, the scripts read in the same files which created repetition when combining scripts. Additionally, some naming conventions were inconsistent across scripts. One notable example was reading in a raw text document but naming it `corpus` when in another file a `quanteda` corpus had been created and also named `corpus` (Rodriguez et al., 2023b). These inconsistencies created some challenges that required us to further investigate the code.

Large file sizes

The files for the comparison of congressional records and parliamentary speeches for the word "empire" were extremely large, with both over 1GB (Rodriguez et al., 2023b). The computation time was long to read these files into R and generally made it more difficult to complete the analysis.

IV. Extension

Validation: Convergent Construct

The authors mention strategies to validate their results, but do not execute any of these validation measures. We investigated the suggestions the authors cited from the 2010 study by Quinn et al. and determined that "convergent construct" validity was the best measure to use to validate case 2 (Quinn et al., 2010). We look first at pure counts of mentions of the word "empire" over time across each of our groups of interest. This plot displays our results and shows that the usage of "empire" starts to shift right at the identified structural shift of 1949. Prior to the break, the U.K. uses "empire" much more than the U.S. does and then around the shift the opposite is true where the U.S. uses "empire" much more than the U.K. This finding is interesting to give context to the results that as time goes on, the use of empire to talk about historical empires decreases.

Furthermore, we investigate the cosine similarities for embeddings of "empire" between the U.S. and U.K. before and after the 1949 structural break. As noted by Quinn et al., 2010, we would anticipate that the cosine similarities would diverge after the structural break since we assume that the focal word is used differently after that time (Quinn et al., 2010). The results are not telling, with both cosine similarities extremely low when comparing the similarities before and after the break.

Sensitivity Testing: `conText` Parameters

The `conText()` function in the author-written package takes a variety of parameters (e.g., the transformation matrix, matrix of pre-trained embeddings, window size, etc.) (Rodriguez, n.d.). Because of the number of user choices, we feel it is important to explore how sensitive the authors' findings are to these parameters. Although this could be done extensively, we demonstrated an example sensitivity test by re-running the Trump vs. Clinton analysis with the `hard_cut` parameter set = TRUE instead of FALSE. In other words, this sets a more strict requirement that all focal word instances considered must be preceded and followed

by at least the window size number of words. While this change cut down of the corpus size in the analysis, the findings were robust to this change (compare this plot in our repo to Figure 4 in the authors' paper).

Future Extensions

Mentioned previously, the authors note that there are validation measures that could have been taken but were not carried out. These methods, semantic and predictive validity, could be helpful to further validate the authors' results through a different lens (Quinn et al., 2010). Outside of validation methods, there are potential extensions for different words of interest, groupings, or time frames. For example, the "empire" case could be extended to represent a more current focal word or to compare the embeddings of the U.S. to another country.

References

Quinn, K.M., Monroe, B.L., Colaresi, M., Crespin, M.H., & and Radev, D.R. (2010). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science* 54(1): 209–28.

Rodriguez, P. L., Spirling, A., & Stewart, B. M. (2023a). Embedding Regression: Models for Context-Specific Description and Inference. American Political Science Review, 117(4), 1255–1274. https://doi.org/10.1017/S0003055422001228

Rodriguez, P. L., Spirling, A., & Stewart, B.M. (2023b). Replication Data for: Embedding Regression: Models for Context-Specific Description and Inference. *Harvard Dataverse*, V1, https://doi.org/10.7910/DVN/NKETXF UNF:6:gBkWkhpPxkGmXEddHggmJQ==\[fileUNF\]

Rodriguez, P. L. (n.d.). prodriguezsosa/conText [R]. Retrieved April 7, 2025, from https://github.com/prodriguezsosa/conText (Original work published 2023)

GitHub repository

Our replication materials are available at https://github.com/mariabartlett/comp-ling-spring-2025-replication-2.