# Appendix

Here we give an overview of why we adopted the statistical approach we did, report details of how we executed that approach, and provide suggested resources for researchers interested in adopting custom contrasts to test specific hypotheses in their own work.

## *Background on contrast specification*

In regression modeling, categorical predictors can be encoded in a number of different ways. The *GoldVarb* software (Sankoff, Tagliamonte, & Smith, 2018) long used in variationist sociolinguistics, for example, encodes categorical predictors using sum contrasts, which assess how each level of the predictor deviates from the grand mean of the predictor’s levels. *R*’s default, on the other hand, is to encode categorical predictors using treatment contrasts, which assess how each level of the predictor deviates from some specified single reference level. These different contrast specifications have consequences for the interpretation of other predictors’ coefficients, including interaction terms; changing the contrast specification changes not only what values are being compared in the model coefficient for the predictor itself, but also what it means to “hold constant” that predictor when evaluating other predictors. For readers interested in deepening their understanding of contrast specification, Schad, Vasishth, Hohenstein, and Kliegl (2020), and Brehm and Alday (2022) provide useful recent overviews aimed at researchers in the language sciences.

The most commonly-used standard contrast schemes, such as sum contrasts, treatment contrasts, and Helmert contrasts, are not well suited to statistical evaluation of the questions we ask in this paper. None of these schemes makes it possible to fit a single model that will test a basic difference (such as an effect of prime variant) in each of two contexts and then also compare those two differences across the contexts. This matters because our research questions have exactly this “difference of differences” structure: we are interested in the appearance of repetitiveness effects within particular contexts as well as the comparison of those repetitiveness effects across those contexts. A common approach to circumventing this issue is to split the data into subsets and fit separate models to each subset, then draw conclusions based on whether the effects of interest are or are not significant in each model. However, that approach requires drawing informal conclusions about comparisons across models without being able to assess whether those comparisons are themselves statistically robust; as Gelman and Stern (2006) pointed out in their paper of the same title, “The difference between ‘significant’ and ‘not significant’ is not itself statistically significant.”

In this study we improve on this common problem by adopting an approach that allows us to test exactly the comparisons we are interested in: custom contrast coding. We do this using the *emmeans* package version 1.7.2 (Lenth et al., 2019) in *R*, which makes it possible to make any and all desired comparisons by extracting and comparing estimated marginal means from a fitted model.

It is important to understand that these custom contrasts are not qualitatively different from the more familiar contrast schemes that come built into programs like *R*. We could use the method we describe below to manually recreate, say, treatment-coded contrasts, and the estimates would be the same as if we had used *R*’s built-in treatment contrasts. The more familiar “standard” contrasts are merely conventions built up around what analyses are made most expedient by our statistical programming tools.

While the *emmeans* package is often used for what are described as post-hoc comparisons, we do not believe this framing (which draws on the language of controlled experiments and ANOVA) is apt for this corpus analysis. Rather, the custom contrasts we construct and assess using *emmeans* are the *primary* analysis of interest and are planned rather than post-hoc. This is the reason that we report only our custom contrast results in the main text.

We also note that the use of the *emmeans* package for post-hoc analysis is sometimes accompanied by the performance of multiplicity corrections; here we do not perform any corrections for multiple comparisons because we report *fewer* critical tests overall than we would have in a more familiar approach testing the full set of levels of each categorical predictor.

## *Details of our statistical analysis*

Following the approach of Muldoon (2019a), we first create a single predictor containing an eight-level encoding of the 2 (*-mian*-primed vs. *-tou*-primed) × 2 (intraspeaker vs. interspeaker) × 2 (similar vs. dissimilar) possibility space cross-tabulating our three categorical predictors of interest. Readers familiar with *GoldVarb* might notice that this is equivalent to the long-familiar approach of creating an “interaction group” to evaluate interactions between predictors. We then fit a mixed-effects logistic regression model containing the eight-level predictor, its interaction with distance (i.e., characters lapsed between the prime and target), and the control predictors of speaker gender, speaker education, and speaker birth year. All categorical predictors are sum-coded (but note that the contrasts used for the 8-level predictor in the original model don’t affect the subsequent evaluation of the custom contrasts). Birth year was centered by subtracting the average birth year from each speaker’s actual birth year; distance between prime and target is log-transformed (with a base of 2) as described in the text. The fitted model at this stage is reported in Table A1. We then create vectors referring to the levels of the critical predictor that we want to compare, then use those vectors to pull out those levels and perform the relevant comparisons using the *contrast()* and *trend()* functions from *emmeans*.

## *Resources for getting started with the* emmeans *package*

For readers who would like to learn about the use of the *emmeans* package, a natural starting point would be the package’s own vignette (Lenth, Love, & Herve, 2021). Beyond or perhaps even before the vignette, we especially recommend a tutorial titled *Getting started with emmeans* from the blog *Very Statisticious* (Muldoon, 2019b). The same blog also provides a helpful tutorial on how the package can be used to set up custom contrasts (*Custom contrasts in emmeans*) (Muldoon, 2019a), which is the source of the approach we followed here. For a slightly more technical approach to custom contrasts that is helpfully aimed at language scientists, we recommend Schad et al. (2020).

Table A1. *GLM results:* target variant (-tou: 0, -mian: 1) ∼ PrimeVar.SpkrMatch.Similarity\* Distance in characters (log2-transformed) + Gender + Birthyear + Education + (1| speaker): *all the categorical predictors are sum-coded*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std.Error | *z* value | Pr(*>*|z|) |
| (Intercept) | 0.30 | 1.41 | 0.22 | 0.83 |
| Intraspeaker.Dissimilar.Mian-primed | -2.34 | 1.69 | -1.38 | 0.17 |
| Intraspeaker.Dissimilar.Tou-primed | -0.57 | 1.55 | -0.37 | 0.72 |
| Intraspeaker.Similar.Mian-primed | 1.32 | 1.44 | 0.91 | 0.36 |
| Intraspeaker.Similar.Tou-primed | -2.19 | 1.43 | -1.54 | 0.12 |
| Interspeaker.Similar.Mian-primed | 1.73 | 1.56 | 1.11 | 0.27 |
| Interspeaker.Dissimilar.Tou-primed | -0.57 | 1.67 | -0.34 | 0.74 |
| Interspeaker.Similar.Tou-primed | -1.99 | 1.54 | -1.29 | 0.20 |
| Log2.CharactersLapsed | -0.18 | 0.18 | -1.03 | 0.30 |
| Gender (Female) | -0.45 | 0.29 | -1.57 | 0.12 |
| Education (HighSchool.Finished) | 0.68 | 0.41 | 1.69 | 0.09 |
| Birthyear (centered) | 0.01 | 0.02 | 0.29 | 0.77 |
| Intraspeaker.Dissimilar.Mian-primed : Log2.CharactersLapsed | 0.30 | 0.23 | 1.35 | 0.18 |
| Intraspeaker.Dissimilar.Tou-primed : Log2.CharactersLapsed | 0.19 | 0.21 | 0.91 | 0.36 |
| Intraspeaker.Similar.Mian-primed : Log2.CharactersLapsed | 0.05 | 0.19 | 0.28 | 0.78 |
| Intraspeaker.Similar.Tou-primed : Log2.CharactersLapsed | 0.33 | 0.19 | 1.75 | 0.08 |
| Interspeaker.Dissimilar.Tou-primed : Log2.CharactersLapsed | 0.19 | 0.22 | 0.86 | 0.39 |
| Interspeaker.Similar.Mian-primed : Log2.CharactersLapsed | -0.12 | 0.21 | -0.58 | 0.56 |
| Interspeaker.Similar.Tou-primed : Log2.CharactersLapsed | 0.28 | 0.21 | 1.37 | 0.17 |

The model has a C-index of 0.91, indicating very good discrimination between the two locative variants. The model correctly predicts the outcome in 83% of the observations (baseline 50%). Condition index of kappa is 81.44, but we believe the high kappa value is attributable to the use of the 8-level predictor—kappa is inflated when the reference level of a multi-level categorical predictor is small.

**References**

Brehm, Laurel & Alday, Phillip M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language* 125:104334.

Gelman, Andrew & Stern, Hal. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician* 60(4):328–31.

Lenth, Russell, Love, Jonathon & Herve, Maxime. (2021). Comparisons and contrasts in emmeans: emmeans package, Version 1.8.4.1. <https://cran.r-project.org/web/packages/emmeans/vignettes/comparisons.html>.

Lenth, Russell, Singmann, Henrik, Love, Jonathon, Buerkner, Paul & Herve, Maxime. (2019). Package ‘emmeans’. <https://github.com/rvlenth/emmeans>

Muldoon, Ariel. (2019a). Custom contrasts in emmeans. [https://aosmith.rbind.io/2019/04/ 15/custom-contrasts-emmeans/.](https://aosmith.rbind.io/2019/04/15/custom-contrasts-emmeans/)

Muldoon, Ariel. (2019b). Getting started with emmeans. [https://aosmith.rbind.io/2019/03/ 25/getting-started-with-emmeans/.](https://aosmith.rbind.io/2019/03/25/getting-started-with-emmeans/)

Sankoff, David, Tagliamonte, Sali & Smith, Eric. (2018). Goldvarb Z: A multivariate analysis application for Macintosh. http://individual.utoronto.ca/tagliamonte/goldvarb.html

Schad, Daniel J., Vasishth, Shravan, Hohenstein, Sven & Kliegl, Reinhold. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of memory and language* 110:104038.