Confirmatory Hypothesis Testing

Erin M. Buchanan

Last Update 2024-04-05

## 0.1 Hypothesis 1

Hypothesis 1 predicted finding semantic facilitation wherein the response latencies for related targets would be faster than unrelated targets. Hypothesis 1 was analyzed by calculating an intercept-only regression model using the *z*-scored priming response latency as the dependent variable. The priming response latency was calculated by taking the average of the unrelated pair *z*-scored response latency minus the average related pair response latency within each item by language. Therefore, values that are positive and greater than zero (e.g., > 0.0001) indicate priming because the related pair had a faster response latency than the unrelated pair. The intercept and its 95% confidence interval represent the grand mean of the priming effect across all languages. Overall priming was = 0.117, = 0.001, 95%CI[0.114, 0.120]. This result supports Hypothesis 1 as the lower limit of the confidence interval is greater than zero (i.e., a directional comparison) from our pre-registration.

This process was repeated for average priming scores calculated without trials that were marked as 2.50 z-score outliers and 3.00 z-score outliers separately. These results were consistent with overall priming: = 0.104, = 0.001, 95%CI[0.114, 0.106], and = 0.107, = 0.001, 95%CI[0.104, 0.109]. Figure 1 denotes the distribution of the average item *z*-score effects, ordered by size of the overall priming effect for each language. The distributions of the priming scores are very similar with long tails and roughly similar shapes (albeit more variance in some languages).

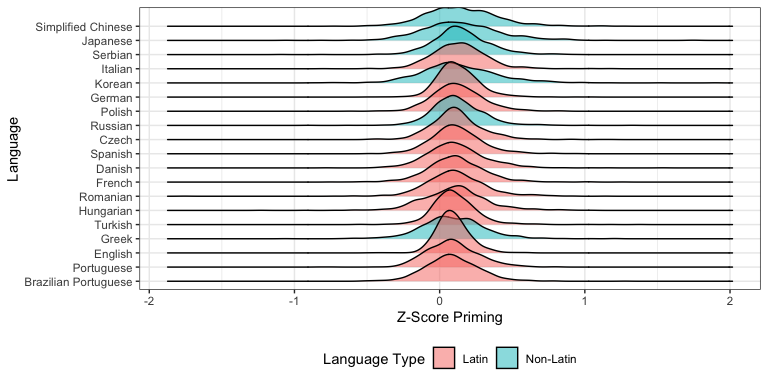


Figure 1: Distribution of average priming effects for languages that met the minimum sample size criteria. Order of languages based on their average priming effect.

## 0.2 Hypothesis 1 - Redefined Accuracy

If we use redefined accuracy as , the results are nearly identical: = 0.116, = 0.001, 95%CI[0.113, 0.119]. This process was repeated for average priming scores calculated without trials that were marked as 2.50 z-score outliers and 3.00 z-score outliers separately. These results were consistent with overall priming: = 0.103, = 0.001, 95%CI[0.113, 0.106], and = 0.106, = 0.001, 95%CI[0.104, 0.109]. Figure 2 denotes the distribution of the average item *z*-score effects, ordered by size of the overall priming effect for each language for the redefined accuracy.

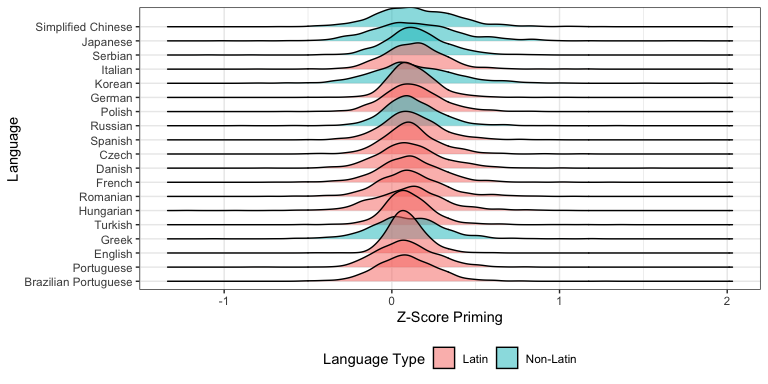


Figure 2: Distribution of average priming effects for languages that met the minimum sample size criteria. Order of languages based on their average priming effect given the redefined accuracy criteria.

## 0.3 Hypothesis 2

Hypothesis 2 explored the extent to which these semantic priming effects vary across languages. Therefore, we calculated a random effects model using the *nlme* (Pinheiro et al. 2017) package in *R* wherein the random intercept of language was added to the overall intercept only model for Hypothesis 1. The addition of this parameter improved model fit from = -6,613.934 to = -6,711.768, supporting significant heterogeneity as the AIC for the random effects model is two points or more less than the AIC for the intercept-only model (Burnham and Anderson 1998). The standard deviation of the random effect was 0.019, 95% CI[0.013, 0.027]. The pseudo- for the model was .008 (Bartoń 2020). The random effect was useful in both *z*-score 2.5 and 3.0 models: = -14,469.541 versus = -14,604.548 and = -12,977.970 versus = -13,104.044. The random effect sizes were similar to the overall model: Z2.5 = 0.017, 95% CI[0.012, 0.024], Z3.0 = 0.017, 95% CI[0.012, 0.025].

Figure 3 portrays the forest plot for the average priming effects by language, ordered by the size of the effect without removal of outliers. The global priming average is presented on each facet to show how the priming effect changes based on the removal of outliers. In nearly all languages, the priming effect decreases slightly with the removal of outliers with the exception of Serbian. This figure also shows that the priming effect does vary by languages, as supported with the results from Hypothesis 2, but that the effect is likely small, given pseudo- was < .01.

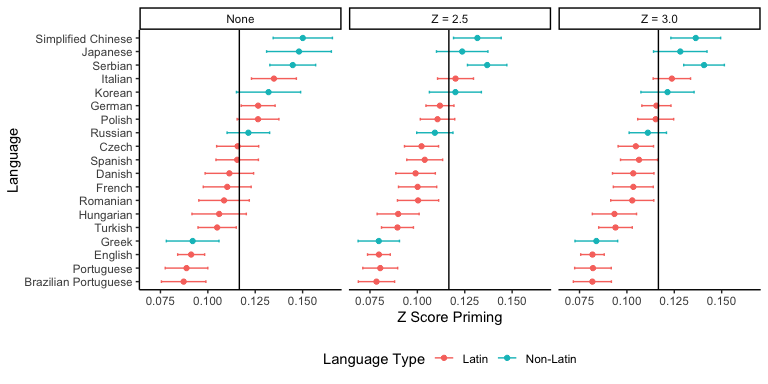


Figure 3: Forest plot of average priming effects for each language ordered by priming average when no outliers are removed. The horizontal line indicates the global average priming effect.

## 0.4 Hypothesis 2 - Redefined

Using the redefined accuracy, we find the same results: the addition of the random intercept of language improved model fit from = -7,005.521 to = -7,097.568. The standard deviation of the random effect was 0.018, 95% CI[0.012, 0.026]. The pseudo- for the model was .008. The random effect was useful in both *z*-score 2.5 and 3.0 models: = -14,865.385 versus = -14,998.100 and = -13,363.540 versus = -13,487.934. The random effect sizes were similar to the overall model: Z2.5 = 0.017, 95% CI[0.012, 0.024], Z3.0 = 0.017, 95% CI[0.012, 0.024]. Figure 4 portrays the forest plot for the average priming effects by language, ordered by the size of the effect without removal of outliers for the redefined effect.

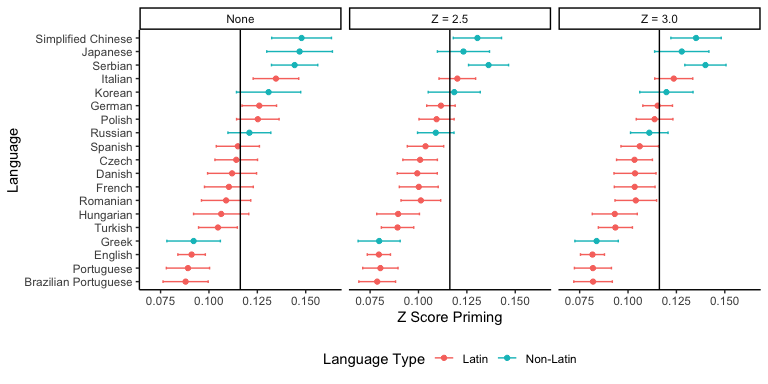


Figure 4: Forest plot of average priming effects for each language ordered by priming average when no outliers are removed for the redefined accuracy results. The horizontal line indicates the global average priming effect.

## References

Bartoń, Kamil. 2020. *MuMIn: Multi-Model Inference*. <https://CRAN.R-project.org/package=MuMIn>.

Burnham, Kenneth P., and David R. Anderson. 1998. “Practical Use of the Information-Theoretic Approach.” In, 75–117. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4757-2917-7_3>.

Pinheiro, J, Douglas Bates, S Debroy, D Sarkar, and R Core Team. 2017. “Nlme: Linear and Nonlinear Mixed Effects Models.” <https://cran.r-project.org/package=nlme>.