

Appendix B: Model construction and selection

Before the initial model (Section 6) was assessed, we first evaluated whether collinearity of all the fixed effects would be an issue. Using the function `collin.fnc()` in the `languageR` library (Baayen, 2013), the condition number (Belsley, Kuh, & Welsch, 2005) was calculated for all the fixed effects, and a value of 4.567 was obtained. Given that the condition number is between 0 and 6, there is very little collinearity between our fixed effects (Baayen, 2008, pp. 198-200). Therefore, we can be confident that each of the fixed effects is a relatively independent component in the regression model.

In constructing our initial model, we did not follow Barr, Levy, Scheepers, and Tily's (2013) recommendation to fit the most complex, convergent random effect structure for our model. This practice has been critiqued elsewhere because it can lead to uninterpretable models (Baayen, Vasishth, Kliegl, & Bates, 2017) with reduced statistical power (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Instead, the random effect structure of the initial superset model was determined by the authors on the basis of the empirical and theoretical considerations discussed in Section 6. The superset model contains all fixed effects, without interaction terms between them, as well as the random intercepts and slopes mentioned above.

(1) Superset model syntax in LME4

Accuracy \sim Stimulus Similarity + Category Similarity (Mean) + Category Similarity (*SD*) + Segmental Frequency + Functional Load + Distributional Overlap + Wordhood + Word Frequency + Neighborhood Density + Average Neighborhood Frequency + Bigram frequency + Response Time + (1 | Unordered Stimulus Pair) + (1 | List) + (1 | Stimulus Order) + (1 | Onset-Coda) + (1 + Segmental Frequency + Functional Load + Segment Environment Distribution + Wordhood + Word Frequency + Neighborhood Density + Average Neighborhood Frequency + Bigram frequency | Participant)

This initial model was then simplified following a step-down, data-driven model selection procedure which compared nested models using the backward best-path algorithm (e.g., Barr et al., 2013; Gorman & Johnson, 2013). We began by simplifying the random effects structure of the model. First, we generated a set of models which were minimally simpler than the superset model, differing only in the omission of one of the random effects. Each of these models was then compared to the superset model using a likelihood ratio test, computed with the `anova()` function. If the likelihood ratio test resulted in a p -value of 0.1 or higher, the simpler model was taken to be an improvement on the superset model.

We chose a relatively liberal threshold of $\alpha = 0.1$ to be conservative in our model selection procedure, preferring to include potentially relevant predictors in the final model if they were reasonably well-justified. In the case that there were multiple subset models which exceeded this α threshold in comparison with the superset model, the subset model with the strongest evidence (the highest p -value) was selected. The random intercepts for both Participant and item (Unordered Stimulus Pair) were never considered for exclusion, as it is standard practice to include these random effects in models of this type (e.g., Jaeger, 2008). This procedure was then repeated, with successive simplification of the random effects structure, until no subset model exceeded the α threshold of 0.1 in comparison with the immediate superset model.

After the best random effect structure was determined, the same steps were repeated to determine the best fixed effects structure. This procedure alternated between random and fixed effect structures until the model could not be reduced any further. The resultant, 'best' model is given in (2).

(2) Best model

Accuracy \sim Stimulus Similarity + Category Similarity (Mean) + Functional Load
+ Distributional Overlap + Word Frequency + (1 | Unordered Stimulus Pair) +
(1 | Participant)

The overall fit of the model, R^2_{GLMM} , was calculated using the `r.squaredGLMM()` function in the `MuMIn` library (Bartoń, 2014; Johnson, 2014; Nakagawa & Schielzeth, 2013). The marginal R^2_{GLMM} is 23.2% and conditional R^2_{GLMM} is 49.9%. Marginal R^2_{GLMM} represents the variance explained by fixed factors and conditional R^2_{GLMM} represents the variance explained by both fixed and random factors. A sizeable portion of the variance (23.2%) was explained with only five fixed factors, suggesting that our relatively simple model is unlikely to be over-fitting the data.

References

- Baayen, R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Baayen, R. (2013). *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=languageR> (R package version 1.4.1)
- Baayen, R., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, 94, 206–234. doi: 10.1016/j.jml.2016.11.006
- Barr, D., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255 - 278. doi: 10.1016/j.jml.2012.11.001
- Bartoń, K. (2014). *Mumin: Multi-model inference* [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=MumIn> (R package version 1.10.0)
- Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). Detecting and assessing collinearity. In M. A. Redford (Ed.), *Regression diagnostic* (pp. 85–191). John Wiley & Sons, Inc. doi: 10.1002/0471725153.ch3
- Gorman, K., & Johnson, D. E. (2013). Quantitative analysis. In R. Bayley, R. Cameron, & C. Lucas (Eds.), *The Oxford handbook of sociolinguistics* (pp. 214–240). Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780199744084.013.0011
- Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434–446. doi: 10.1016/j.jml.2007.11.007
- Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, n/a–n/a. Retrieved from <http://dx.doi.org/10.1111/2041-210X.12225> doi: 10.1111/2041-210X.12225
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, R., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining r^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. doi: 10.1111/j.2041-210x.2012.00261.x