Cognateness, frequency, and vocabulary size: an interactive account of bilingual lexical acquisition

G. Garcia-Castro, D. S. Avila-Varela, I. Castillejo, N. Sebastian-Galles

One of the most prominent features of the bilingual mental lexicon is that the activation of representations spreads in a language non-selective fashion. A clear example of such non-selectivity is embodied by cognateness (i.e., form-similarity between translation equivalents), which impacts lexical access and processing in bilinguals. Recent studies have suggested that cognateness facilitates vocabulary acquisition in bilingual toddlers, who show larger vocabulary sizes when their languages share many cognates (Floccia et al., 2018), and who also acquire cognates earlier in age than non-cognates (Mitchell et al., 2022). The specific mechanisms underpinning such facilitation effect are unclear. In this study, we present a model of bilingual early lexical acquisition in which cognateness interacts with lexical frequency and language exposure to facilitate the acquisition of low-frequency words. We tested this model against vocabulary data from 436 Catalan-Spanish bilinguals aged 12 to 34 months. We used Bayesian Item Response Theory to estimate participants' probability of acquisition of 604 words, conditional to the cognate status and lexical frequency of the word-form, and the age and degree of exposure to each language of the toddler. We found converging evidence for an earlier age of acquisition for cognate words, and for such effect being mediated by lexical frequency and language exposure. Low-frequency words, and words from the language of least exposure benefited more strongly by their cognate status than high-frequency words. Overall, our findings support an interactive account of bilingual vocabulary acquisition in which the lexical representations in one language interact with the acquisition and processing of words in the other language.

References

Floccia, C., Sambrook, T. D., Luche, C. D., Kwok, R., Goslin, J., White, L., Cattani, A., Sullivan, E., Abbot-Smith, K., Krott, A., Mills, D., Rowland, C., Gervain, J., & Plunkett, K. (2018). I: Introduction. *Monographs of the Society for Research in Child Development*, *83*(1), 7–29. https://doi.org/10.1111/mono.12348

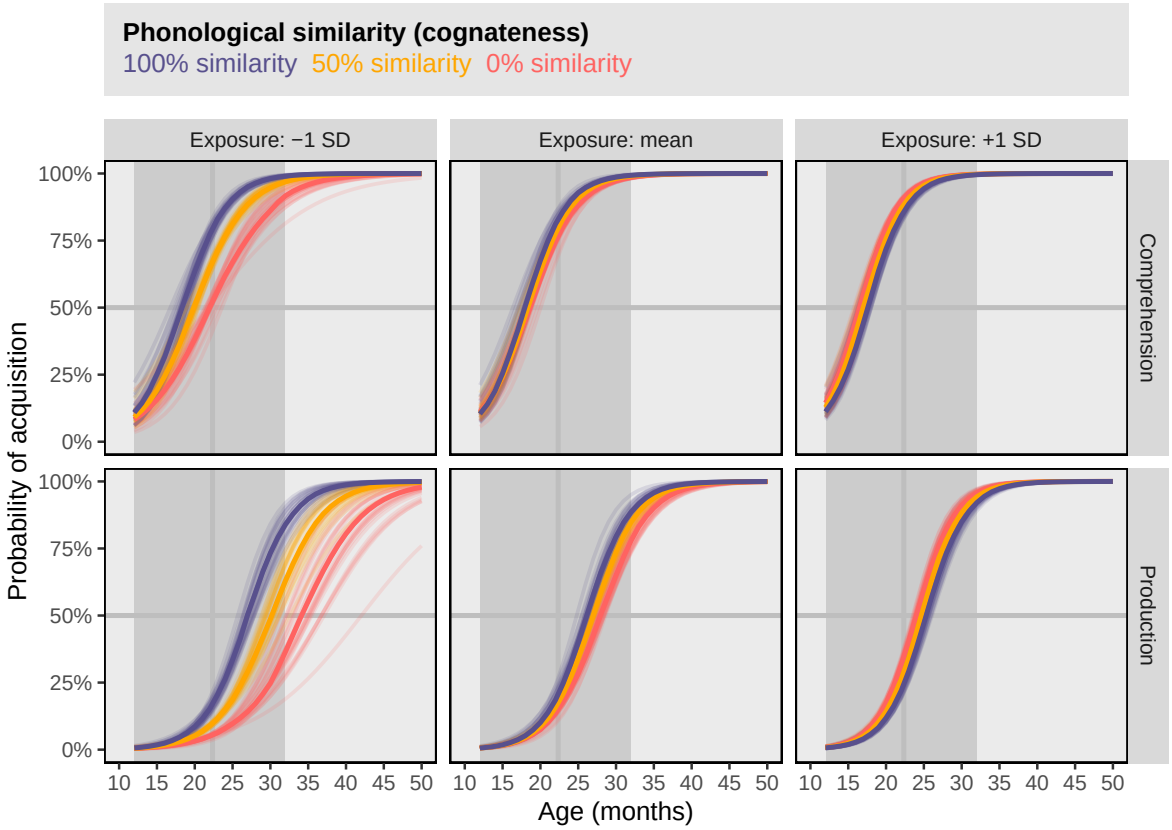

Figure 1: Expected posterior predictions from the multilevel ordinal regression Bayesian model. We show 25 random posterior predictions, each for a series of combination of levels of interest. Each thin line and corresponds to one prediction. Thicker lines indicate the median of the predictions. The X-axis indicates the age (in months) for which the prediction is generated. The Y-axis indicates the predicted probability of acquisition (*Comprehension* or *Comprehension and Production*). Different colours indicate different levels of phonological similarity, operationalised as the Levenshtein similarity between pairs of translation equivalents. Predictions are presented separately for different degrees of word exposure index: little exposure to the word (-1 SD), median exposure, and high exposure (+1 SD). Predictions for *Comprehension* are show on top and predictions for *Comprehension and Production* are shown on the bottom. In-sample predictions lie inside the grey rectangles.

Mitchell, L., Tsui, R. K. Y., & Byers-Heinlein, K. (2022). *Cognates are advantaged in early bilingual expressive vocabulary development.* PsyArXiv. https://doi.org/10.31234/osf.io/daktp