

Developmental Trajectories of Vocabulary Composition Across Languages

Mika Braginsky, Daniel Yurovsky, Virginia Marchman, and Michael C. Frank

More comprehensive look at vocabulary composition – many languages, large samples of children, over development.

- Most languages show a positive noun bias: children produce more nouns than would be expected by chance.
- Extent of this bias varies cross-linguistically, with Mandarin and Cantonese at the low end of the continuum.
- Degree of noun and predicate bias are strongly negatively correlated ($r=-0.74$).
- Across all languages, function words are substantially underrepresented, although there is considerable variability in the degree of this negative bias.

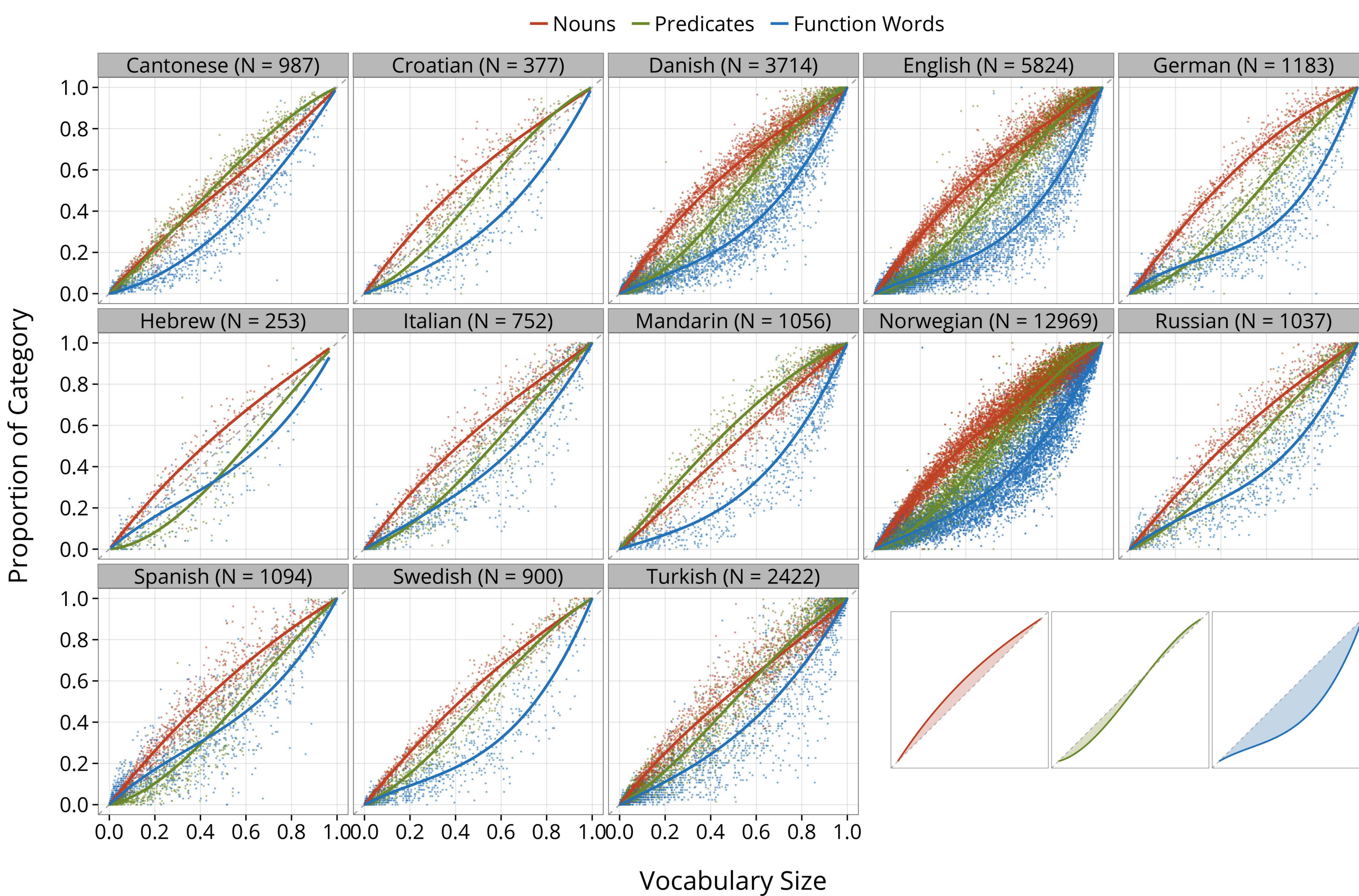
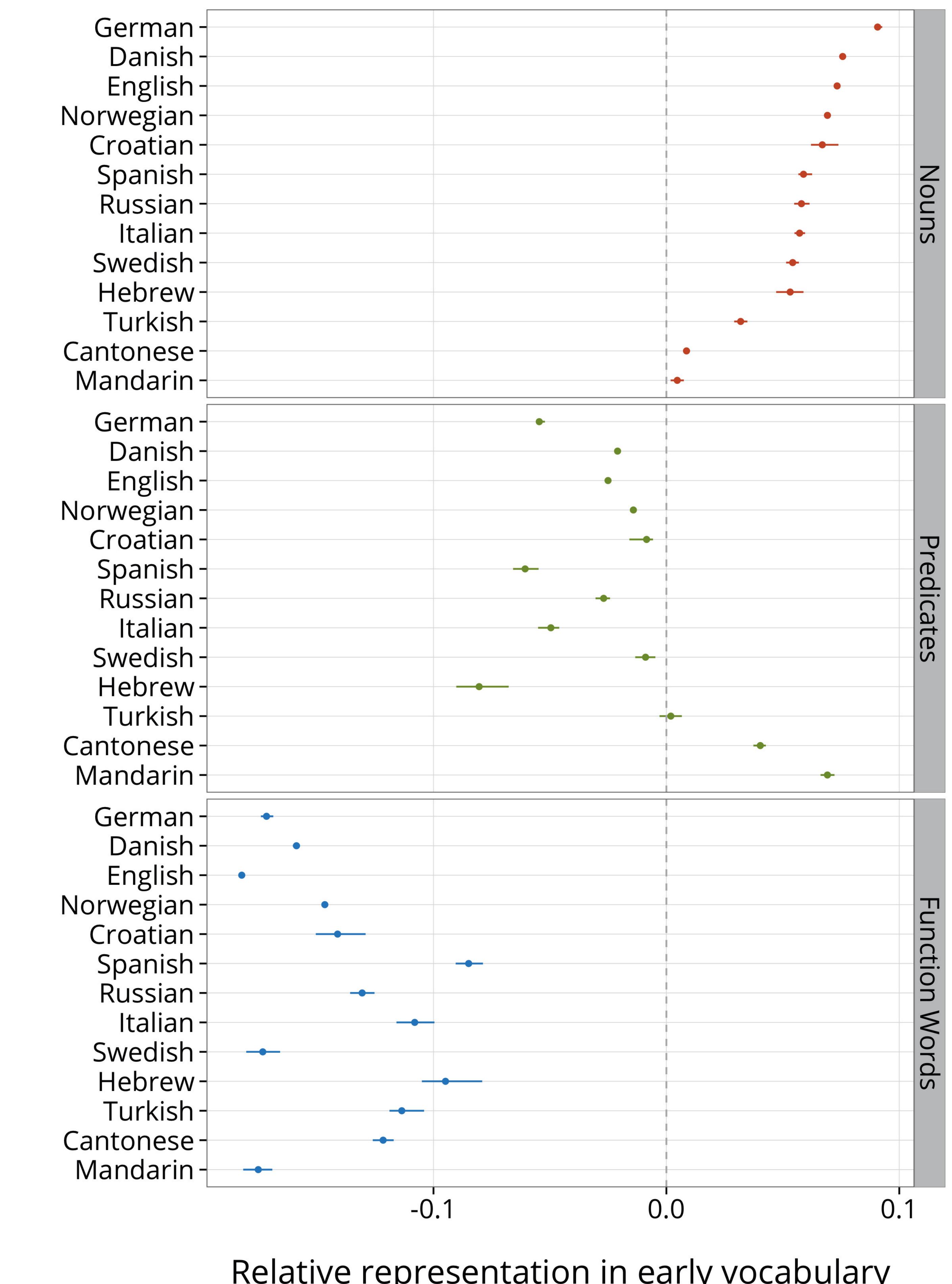


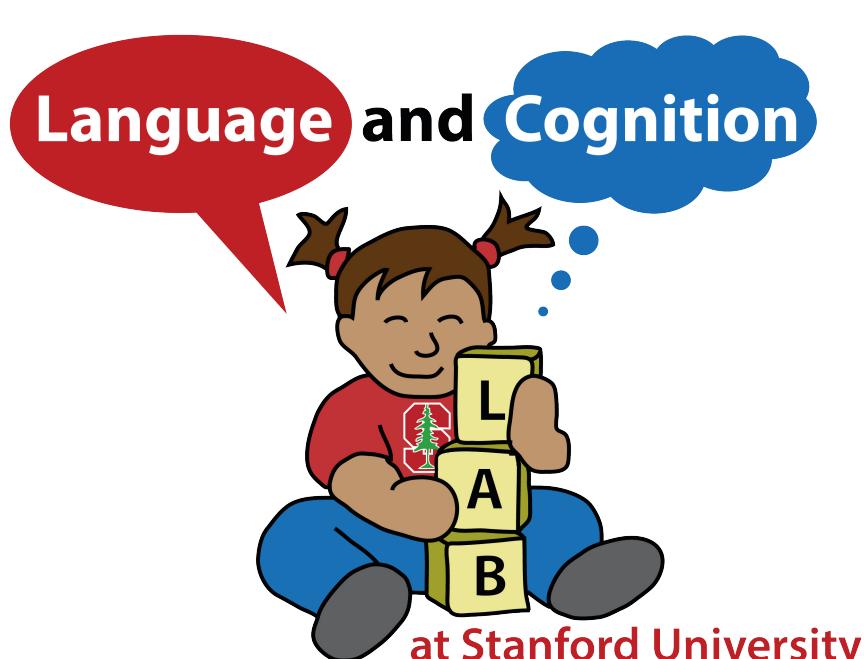
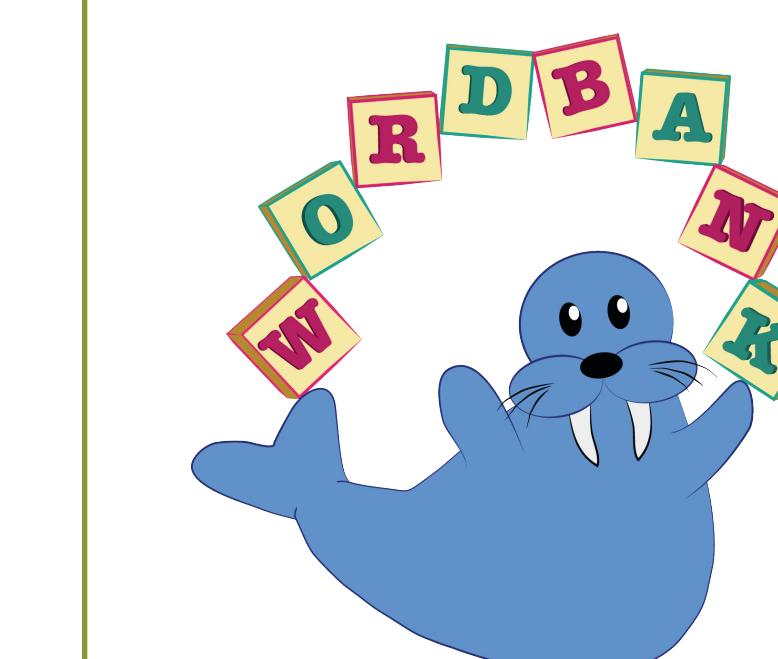
Figure 1: Each MB-CDI administration is plotted as a point reflecting the proportion of nouns, predicates, and function words on the form that the child is reported to produce as a function of the proportion of all three combined that the child is reported to produce; lines indicate constrained polynomial curves fit to those data.



Relative representation in early vocabulary

Figure 2: Bias estimates (area between the model predictions and the diagonal) for each language and lexical category, with line segments showing bootstrapped 95% confidence intervals.

Data and code for these analyses are available at
<https://github.com/mikabr/vocab-comp>



wordbank.stanford.edu