

Consistency and variability in word learning across languages

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

- Why do children learn some words earlier than others?
- Factors contributing to word difficulty can provide clues about the mechanisms of word learning
- Estimate the acquisition trajectories for around 400 words in 10 languages
- Predict them on the basis of independently-derived environmental and conceptual factors
- Examine consistency and variability of these predictors across languages, by lexical category, and over development

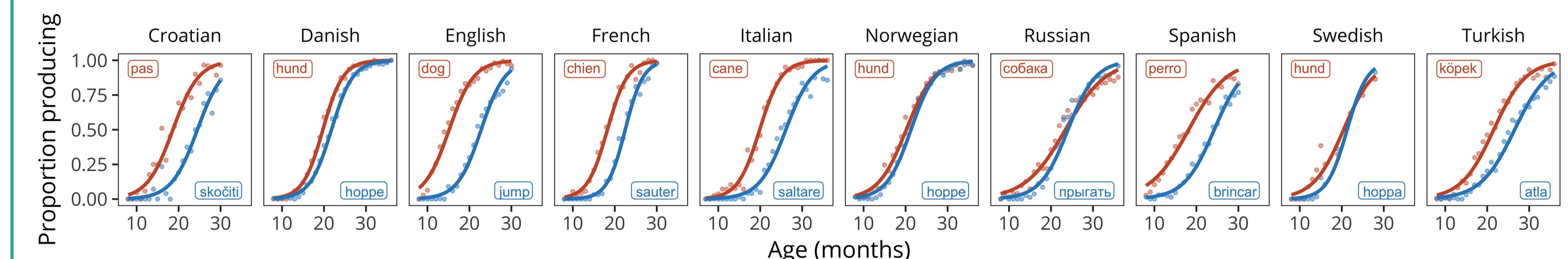


Figure 1: Example production trajectories for the words "dog" and "jump" across languages. Points show average proportion of children producing each word for each one-month age group. Lines show the best-fitting logistic curve. Labels show the forms of the word in each language.

Language	CDI Items	N production	N comprehension	Predictor	Highest	Lowest
Croatian	390	627	250	Babiness	baby, bib, bottle	donkey, penny, jeans
Danish	383	6112	2398	MLU	when, day, store	peekaboo, ouch, hello
English	393	5967	1821	Frequency	you, it, that	cockadoodledoo, grrr, church
French	396	1364	537	Concreteness	apple, baby, ball	how, now, that
Italian	396	1400	648	Solo frequency	no, yes, what	tooth, feed, aunt
Norwegian	381	7466	2374	Arousal	naughty, money, scared	shh, asleep, blanket
Russian	410	1805	768	Length	cockadoodledoo, refrigerator, rocking chair	i, go, hi
Spanish	399	1872	778	Valence	happy, hug, love	sick, hurt, ouch
Swedish	371	1367	467	Final frequency	book, it, there	give, when, put
Turkish	396	3537	1115			

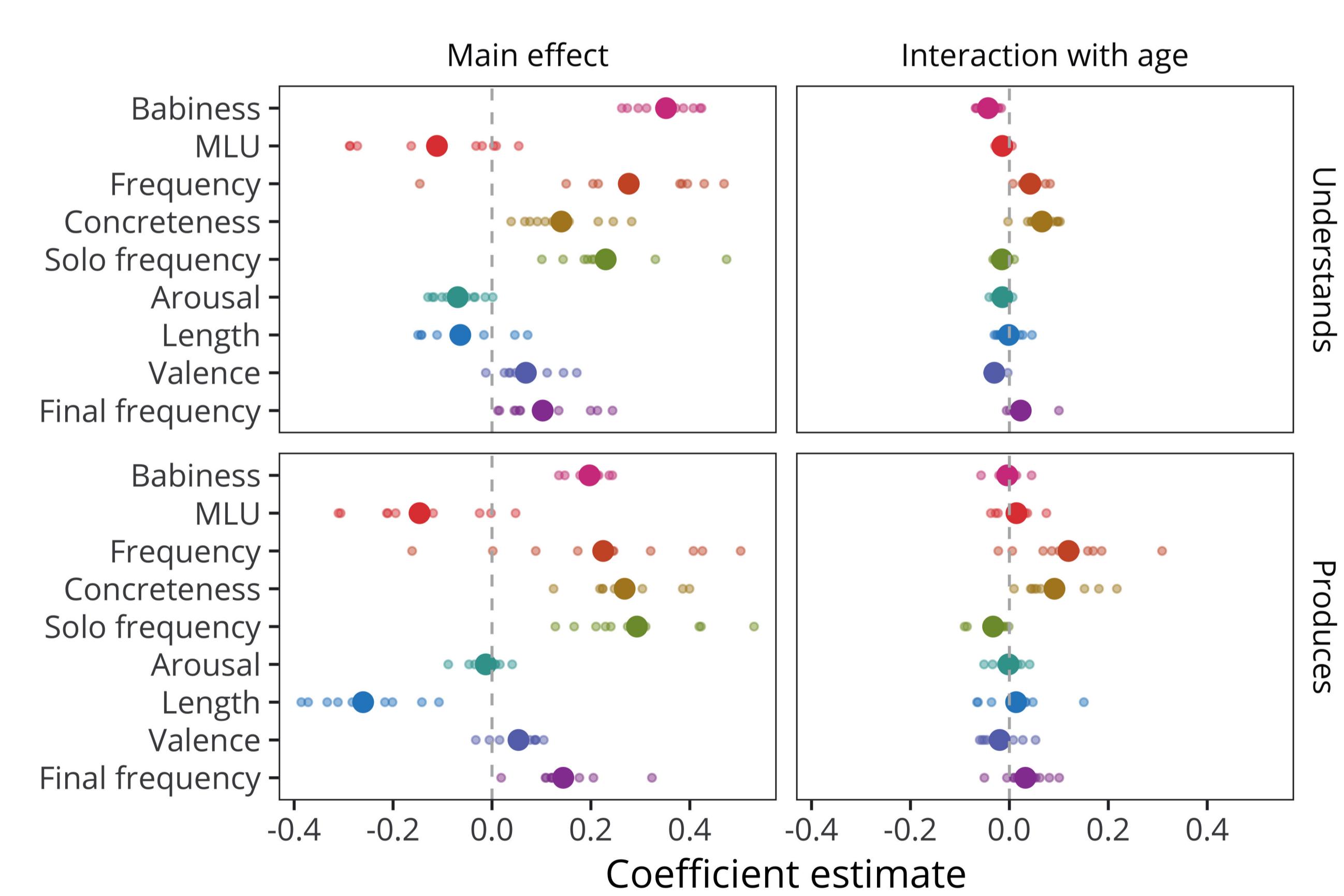


Figure 2: Estimates of coefficients in predicting words' developmental trajectories. Each point represents a predictor's coefficient in one language, with the large point showing the mean across languages. Larger coefficient values indicate a greater effect of the predictor on acquisition: positive main effects indicate that words with higher values of the predictor tend to be understood/produced by more children, while negative main effects indicate that words with lower values of the predictor tend to be understood/produced by more children; positive age interactions indicate that the predictor's effect increases with age, while negative age interactions indicate that the predictor's effect decreases with age.

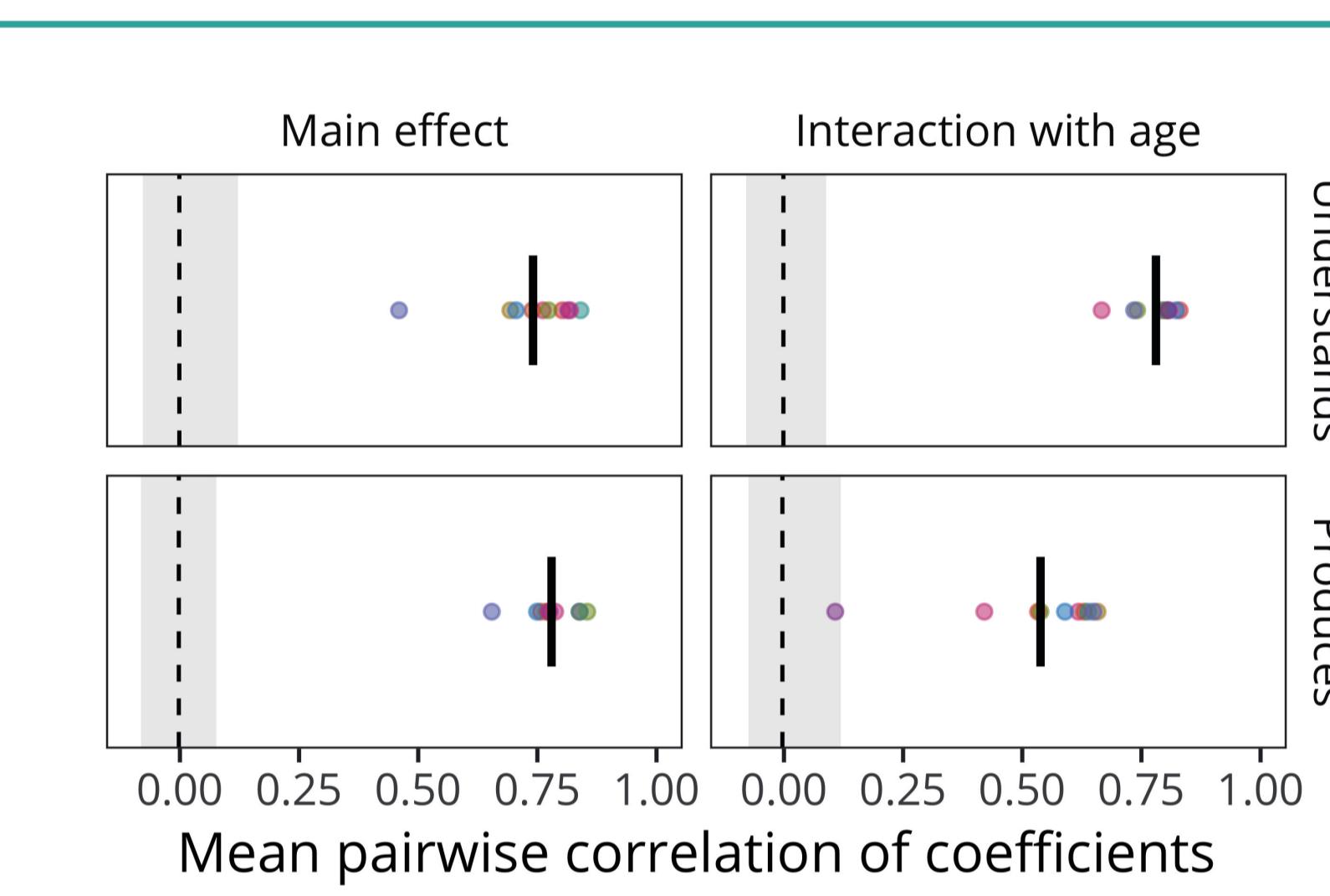


Figure 3: Correlations of coefficients estimates between languages. Each point represents the mean of one language's coefficients' correlation with each other language's coefficients, with the black line indicating the overall mean across languages. The grey region and dashed line show a bootstrapped 95% confidence interval of a randomized baseline where predictor coefficients are shuffled within language.

- Babiness, frequency, MLU, and concreteness are relatively stronger predictors
- Considerable consistency in the magnitudes of predictors across languages
- Word length far more predictive for production than comprehension
- Effects of concreteness and frequency increase with age
- Effects of babiness and valence (comprehension) and solo frequency (production) decrease with age

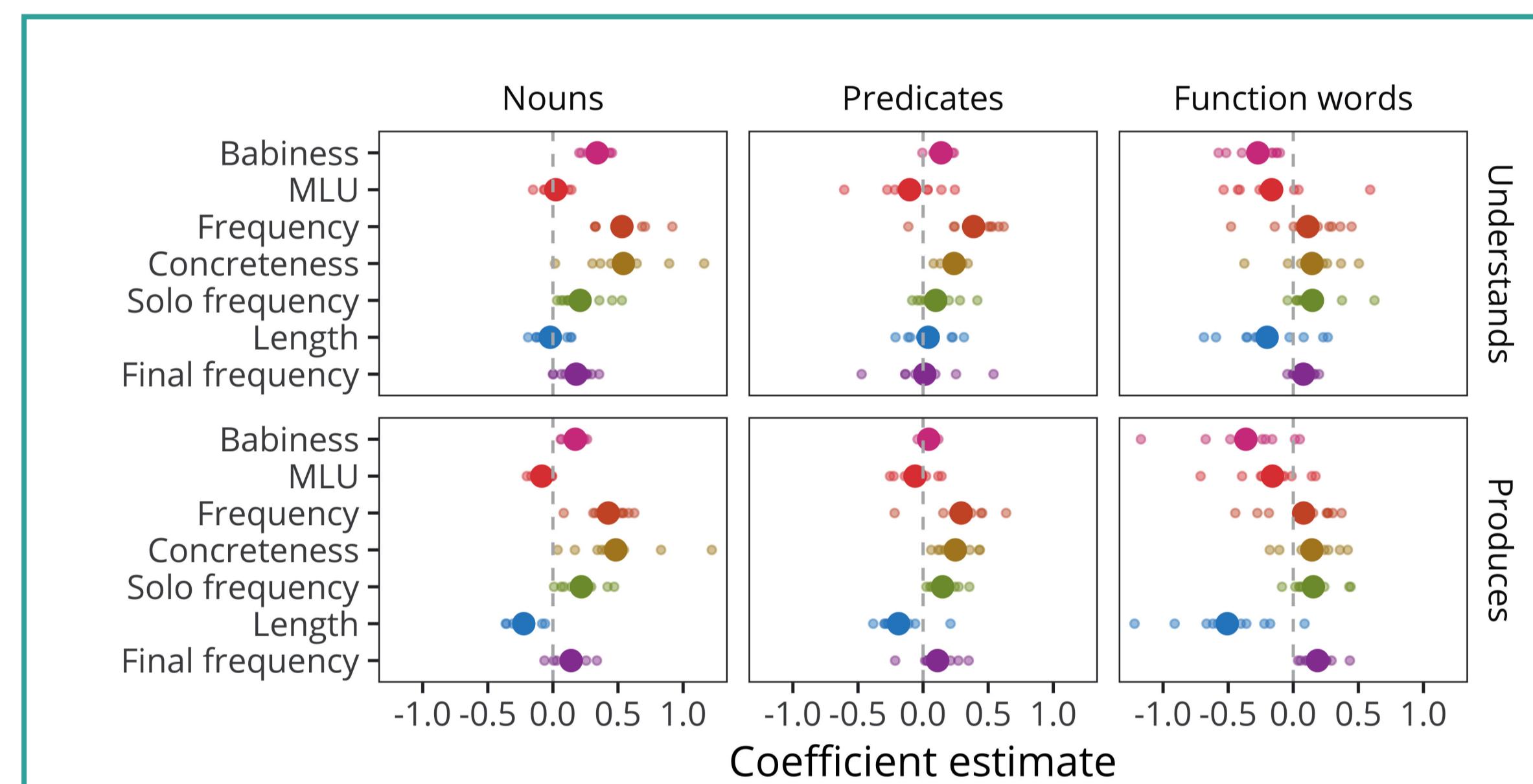
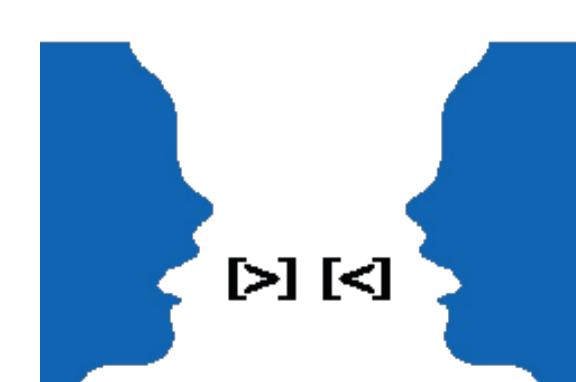


Figure 4: Estimates of coefficients in predicting words' developmental trajectories (as described in Figure 2), with separate models for each lexical category.

- Frequency highest for nouns, lower for function words
- MLU irrelevant for nouns and predicates, highly predictive for function words
- Supports hypothesis that different word classes are learned in different ways

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



wordbank.stanford.edu