

## Gender Differences in Lexical Input and Acquisition

Throughout childhood, children are treated differently based on their assigned gender, disseminating culturally-determined gender roles and influencing their self-perceptions and social behaviors (Eccles et al., 1993; Eckes & Trautner, 2000). These socialization practices should be reflected in the language that parents use to children, which in turn affects language acquisition itself. To investigate this interconnection, we analyzed gender-based lexical differences in input, acquisition, and their relationship. In a large corpus of child-directed speech from CHILDES (MacWhinney, 2000), we found that many words show frequency differences by children's gender. We also measured word-level gender differences in large-scale vocabulary acquisition data using parent-reported Communicative Development Inventory (CDI) measures (Fenson, 2007). Following the approach of predicting words' learnability from independently measured properties (Goodman, Dale, & Li, 2008; Roy et al., 2015), we found that gender differences in input frequency are predictive of gender differences in acquisition trajectory.

Looking at language input, we examined how the lexical properties of child-directed speech vary by gender. Our dataset included all adult speech in the American English portions of CHILDES with a single target child of known gender. For each word in the corpus mappable to a CDI item, we computed the log ratio between its probabilities in girl-directed and boy-directed speech (Figure 1). We used the log-likelihood ratio test (Dunning, 1993) to compare the probability of each word's counts under the gender-separated corpora and the joint corpus: 61% (375/610) of words' frequency differences significantly deviated from chance ( $p < 0.01$ ); with the Bonferroni correction for multiple comparisons, 41% (251/610) were significant.

For learning outcomes, we characterized differences in vocabulary acquisition using CDI data from Wordbank (Frank et al., in press). For each word on Words and Gestures (WG; 8-18 months) and Words and Sentences (WS; 16-30 months), we fit logistics to the acquisition trajectories of girls and boys and estimated the area between them (Figure 2). This area difference measures how many more children of one group understand/produce a word over development. Girls understood and produced substantially more words than boys overall. Words' gender-based differences were highest among concrete nouns: the most girl-biased categories were clothing and body parts, while the most boy-biased category was vehicles.

Finally, we analyzed the relationship between input frequency and acquisition. We fit a linear regression to predict area difference from log probability ratio in girl-directed and boy-directed speech (Figure 3), finding a significant effect of log probability ratio for all three measures (WS produces and WG understands  $p < 1e-10$ ; WG produces  $p < 0.01$ ). Although the model accounted for only a modest amount of the variance in area difference, many factors besides input frequency are relevant to a word's learnability: the predictive relationship suggests that to the extent that frequency relates to acquisition, its impact is differentiated by gender.

In summary, we investigated the extent to which signatures of gender-based socialization processes can be found in properties of children's lexical input and acquisition, and found strong signals differences by assigned gender in the lexical statistics of input, learning outcomes, and the relationship between them.

(499 words)

## References

- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*.
- Eccles, J., Wigfield, A., Harold, R. D. and Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development*.
- Eckes, T., & Trautner, H. M. (2000). The developmental social psychology of gender. Psychology Press.
- Fenson, L. (2007). MacArthur-Bates Communicative Development Inventories: User's guide and technical manual. Paul H. Brookes Publishing Company.
- Frank, M. C., Braginsky, M., Yurovsky, D., Marchman, V. A. (in press). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*.
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*.
- MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2: The Database. Mahwah, NJ: Lawrence Erlbaum Associates.
- Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*.

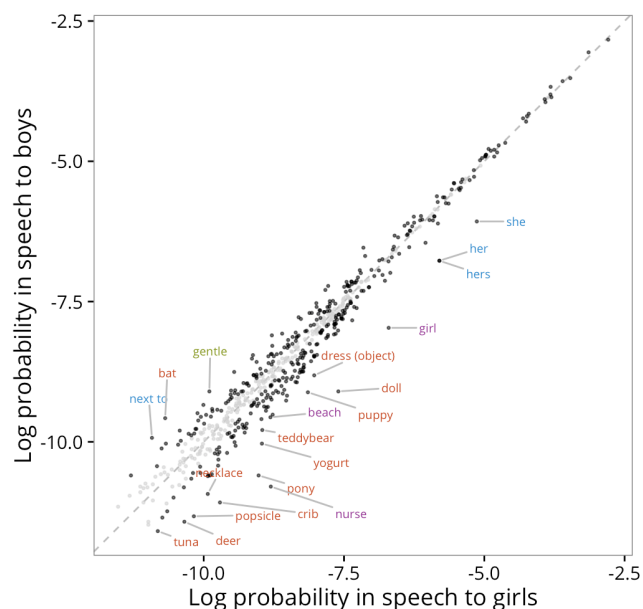


Figure 1. Frequency differences in the input (black points indicate words whose frequencies are significantly different by log likelihood ratio test; top 20 words are labeled).

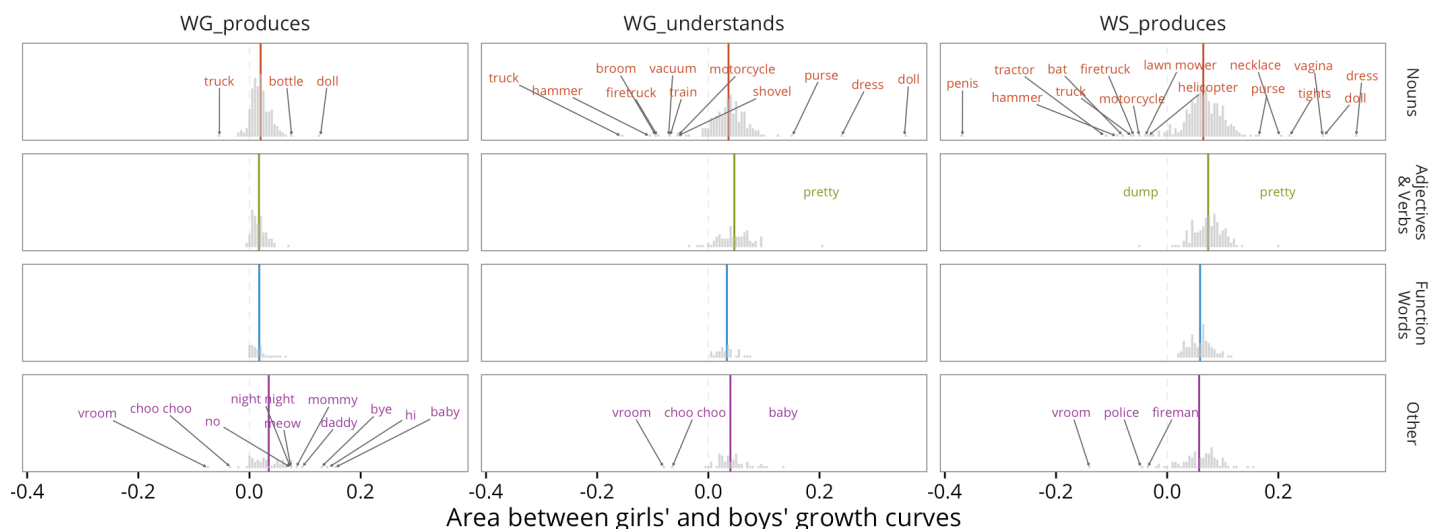


Figure 2: Girls learn most words earlier than boys, with item-specific differences (labeled words are in the 99<sup>th</sup> percentile).

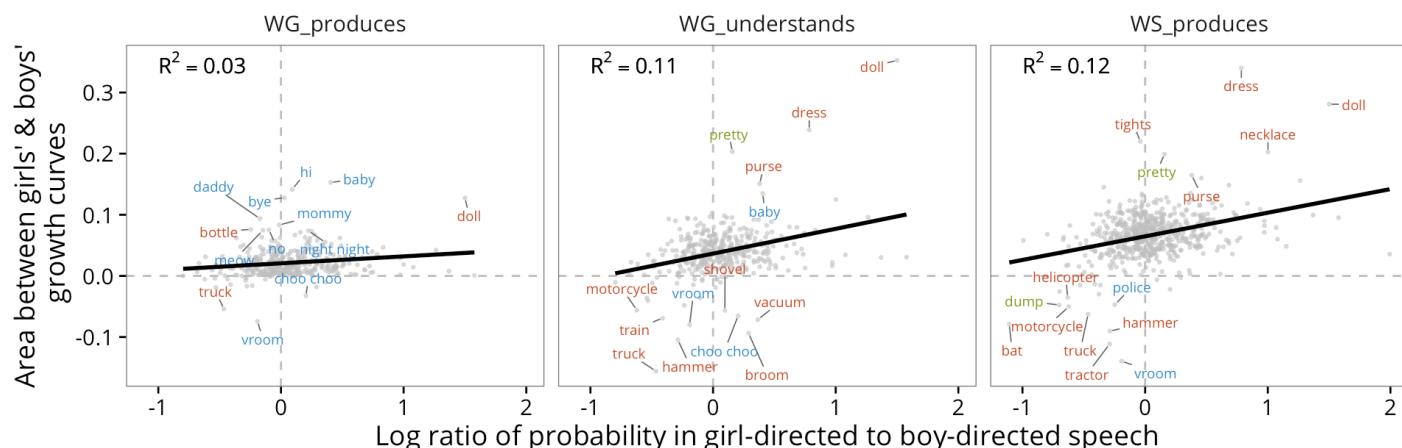


Figure 3: Input gender differences predict early production and comprehension measures (labeled words are in the 99<sup>th</sup> percentile).