

# Children's books as a unique source of early language input

Joseph Denby & Daniel Yurovsky  
The Department of Psychology, University of Chicago



## Background

A child's language development largely depends on their early linguistic environment. For decades, quality and quantity of child-directed speech are shown to be strongly predictive of language maturity.

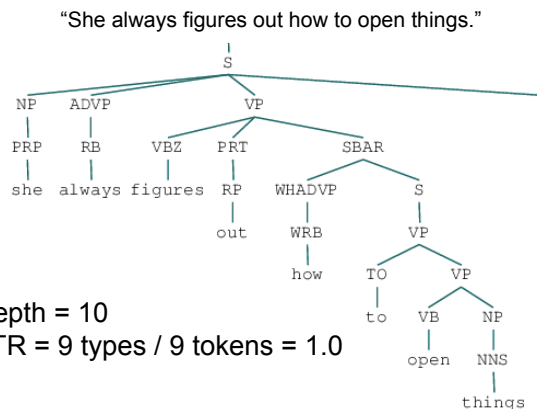
But, picture books are another prominent part of early linguistic input. How are they different from caregiver speech? Montag et al., (2015) finds that book text is more lexically diverse than caregiver speech (using the CHILDES corpus):

**To what extent can books supplement a child's early linguistic environment?**

## Corpora / Analyses

- **Language Development Project** (Goldin-Meadow et al., 2015)
  - Parent-Child conversations ( $n = 102$ ; ages 14-58 months)
  - Socioeconomically diverse sample of families
- **Children's Book Corpus** (Montag et al., 2015)
  - 100 popular books sourced from librarian recommendation, Amazon best-seller lists, and library circulation statistics

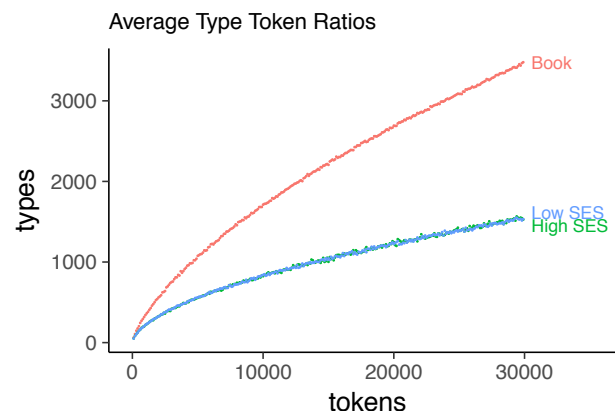
To evaluate lexical diversity, sentence complexity, and specific language usage, we employed type token ratio computation, sentence parsing trees, and part-of-speech tagging.



## Book Language is More Diverse

**Children's books are more lexically diverse than child-directed speech**

- Evaluated with type-token ratios on random length-matched samples

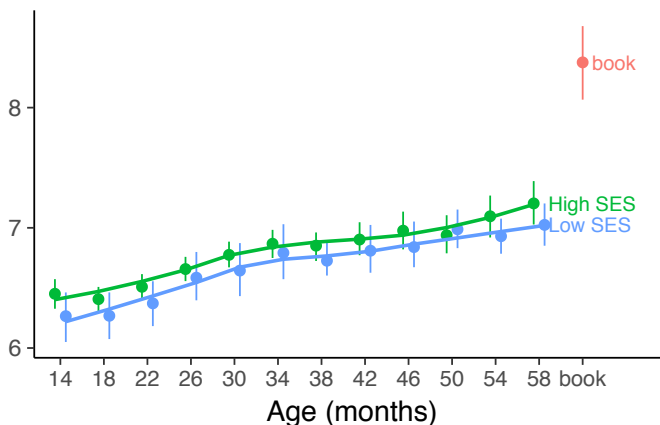


## Book Language is More Complex

**Children's books have a higher average sentence complexity than child-directed speech**

- Using parse tree depth as a complexity proxy on random length-matched samples

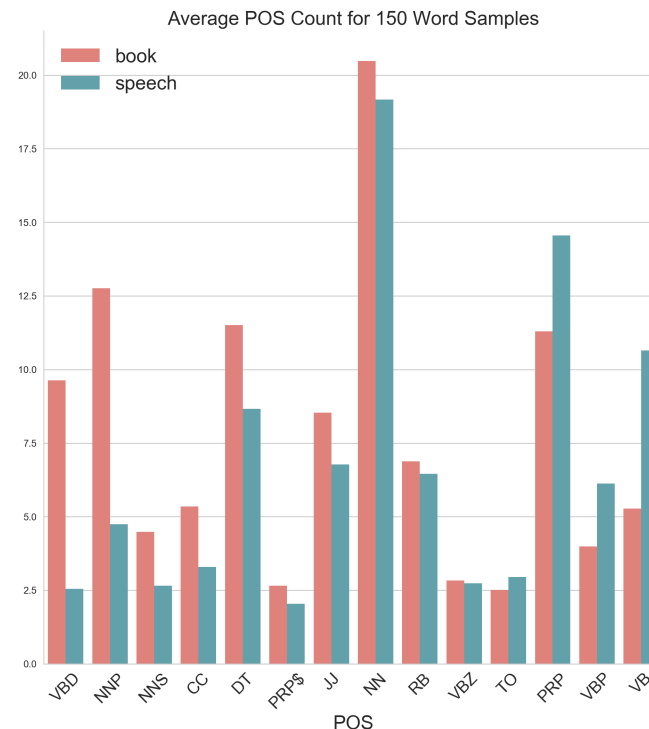
Average Sentence Complexities



## Book Language Is Unique

**Children's books use parts-of-speech uniquely when compared to child-directed speech**

- Tagging using Penn-Treebank tagset and nltk



## Conclusions

- Children's book text is markedly different from caregiver child-directed speech along several dimensions
- Some of these dimensions, within speech, are positively correlated with development outcomes (PPVT, CDI)
- Books are well-poised to supplement a child's early linguistic environment by making certain kinds of speech more available.