

# The Words Children Hear: Picture Books and the Statistics for Language Learning

Psychological Science

1–8

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0956797615594361

pss.sagepub.com



Jessica L. Montag, Michael N. Jones, and Linda B. Smith

<sup>1</sup>Department of Psychological and Brain Sciences, Indiana University Bloomington

## Abstract

Young children learn language from the speech they hear. Previous work suggests that greater statistical diversity of words and of linguistic contexts is associated with better language outcomes. One potential source of lexical diversity is the text of picture books that caregivers read aloud to children. Many parents begin reading to their children shortly after birth, so this is potentially an important source of linguistic input for many children. We constructed a corpus of 100 children's picture books and compared word type and token counts in that sample and a matched sample of child-directed speech. Overall, the picture books contained more unique word types than the child-directed speech. Further, individual picture books generally contained more unique word types than length-matched, child-directed conversations. The text of picture books may be an important source of vocabulary for young children, and these findings suggest a mechanism that underlies the language benefits associated with reading to children.

## Keywords

language development, statistical analysis, reading, computer simulation

Received 3/5/15; Revision accepted 6/12/15

The talk that surrounds human infants provides data for language learning. A large literature indicates that talk directed to the child—rather than adult-adult or background talk—is the core data on which early language learning depends (e.g., Weisleder & Fernald, 2014). Therefore, studies of the input relevant to early language learning have focused on conversations between parents and children. Major advances have emerged from analyses of the statistical properties of large corpora aggregated over many such conversations (see Hills, Maouene, Riordan, & Smith, 2010; MacWhinney, 2000; Ninio, 2011).

Most relevant to the present research, computational analyses of the child-directed language in these aggregated corpora indicate a key role for the diversity of the input; learnability is enhanced when individual words or linguistic devices are encountered in diverse contexts and when there is more variety in general in the input data (Hills et al., 2010; Jones, Johns, & Recchia, 2012). Studies of individual differences in child-directed language also point to diversity in parents' talk as a critical predictor of the outcome of language learning (Hart & Risley, 1995; Hoff, 2003; Hoff & Naigles, 2002; Huttenlocher, Waterfall,

Vasilyeva, Vevea, & Hedges, 2010; Pan, Rowe, Singer & Snow, 2005; Rowe, 2012; Weizman & Snow, 2001). Here we focus on a possible source of lexical diversity in early learning environments, the text in children's picture books. This is a source that is associated with improved language outcomes for children (i.e., children who are read to more often by caregivers exhibit better language skills; Farrant & Zubrick, 2012; Payne, Whitehurst, & Angell, 1994; Sénéchal & LeFevre, 2002), but one that has not been systematically studied for its statistical properties.

Infants do not read picture books, but many parents begin regularly reading picture books to their infants shortly after birth (Deckner, Adamson & Bakeman, 2006; Karrass & Braungart-Rieker, 2005; Young, Davis, Schoen & Parker, 1998). Large representative surveys of parents indicate that more than 50% of parents of infants ages

## Corresponding Author:

Jessica L. Montag, Department of Psychology, University of California, Riverside, 900 University Ave., Riverside, CA 92521  
E-mail: jessica.montag@ucr.edu

0 to 5 months report reading books to their infant at least once a week and more than 25% of parents of infants ages 6 to 11 months report reading to their infant at least once a day (Young et al., 1998). Other studies indicate that parents both chat conversationally about the contents of books with their children and read the text as written (Deckner et al., 2006; Dickinson, Griffith, Golinkoff, & Hirsh-Pasek, 2012; Fletcher, Cross, Tanney, Schneider, & Finch, 2008; Mol, Bus, de Jong, & Smeets, 2008; Ninio & Bruner, 1978; Whitehurst et al., 1988). Thus, the words in child-directed picture books are part of parents' child-directed talk.

Analyses indicate that there are marked differences in vocabulary choice between adult-directed conversations and adult-directed written texts. In an analysis of everyday conversations between adults and a variety of written-text samples, conversations were found to use relatively few words (93.9% of all words were contained in a 5,000-word basic lexicon), whereas the text samples included a more diverse set of words (84.3% and 88.4% of the words in the newspapers and books, respectively, were contained in the basic lexicon). This work suggests that there are important differences in the lexical diversity of written and spoken language (Hayes & Ahrens, 1988). If vocabulary differences between child-directed speech and the text in picture books that parents read to young children mirror these differences in lexical diversity observed in adult-directed language, then the language in picture books may play a significant, though as yet unstudied, role in early language learning. Further, and as we consider in the General Discussion, if these early books do present different data for learning than parent-child conversations do, then individual differences in shared book reading may contribute substantially to the well-documented individual differences in early language learning and their far-reaching consequences for later language processing (Sénéchal & LeFevre, 2002).

Accordingly, we asked: What language-learning data might early picture books provide that everyday conversations do not? We addressed this question by comparing the lexical diversity in parent-child conversations and in the texts of picture books. Our principle measure of lexical diversity was the number of unique words (types) relative to the total number of words (tokens). For example, the phrase *the cat and the dog* has a total of five word tokens, because the phrase is five words long, and four word types, because *the* is repeated, so that there are four unique words. Type and token counts and type-token ratios are widely used in the study of language development, and individual differences in type-token ratios in the language-learning environment are predictive of vocabulary development (Huttenlocher et al., 2010; Pan et al., 2005; Rowe, 2012; Weizman & Snow, 2001).

## Method

### *The corpora*

**Words in children's picture books.** There is no existing corpus of children's picture books, so we constructed a corpus for our analyses. Our corpus consisted of the text of 100 children's picture books (68,103 words) that a caregiver might read to infants and very young language learners. In order to obtain a sample that is representative of the books that parents read to very young children, we selected the titles from lists of librarian-recommended picture books, amazon.com best sellers, and circulation statistics from the Infant and Preschool sections of the Monroe County (Indiana) Public Library. These are books recommended to parents and read by parents of infants and children ages 0 to 60 months, and thus within the age range of the conversational CHILDES (Child Language Data Exchange System) corpora. The list of books is provided in Table 1.

**Words in child-directed conversations.** We obtained our sample of child-directed speech from the American English subset of the CHILDES corpus (MacWhinney, 2000). We limited our sample to speech to children ages 0 to 60 months to match the age range to the intended age range of the picture books. Our sample comprised 4,432 individual conversations (contiguous recording sessions) across a variety of settings, for a total of about 6.5 million words of speech. We used a version of the CHILDES corpus that had been processed to (a) remove a number of the special transcription characters and other artifacts of the CHILDES coding system and (b) systematize words with idiosyncratic spellings (e.g. replace all instances of "doggy" with "doggie" to maintain consistent spelling; Willits & Jones, 2015).

### *Sampling procedure*

Type-token ratios depend on sample size and therefore cannot be easily interpreted using a single sample or measure (Hess, Sefton, & Landry, 1986; Richards, 1987). Accordingly, in our analyses, we focused on the number of unique word types found at multiple sample (token) sizes. To obtain a distribution of type counts as a function of tokens, we randomly sampled word sets of various sizes from the two corpora.

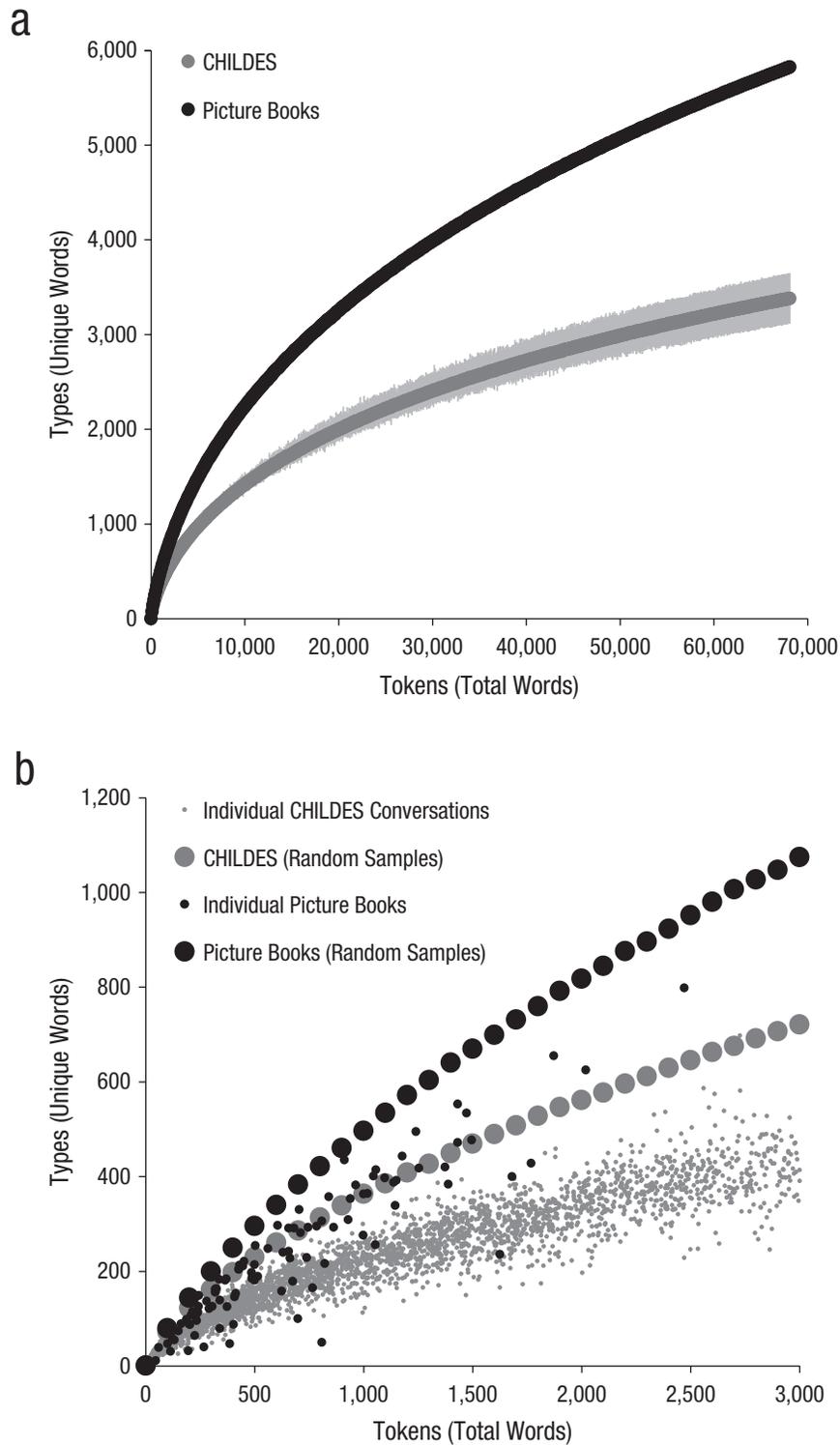
**Picture books.** The different-sized random samples from the picture books were constructed by taking progressively larger random samples that increased in increments of 100 words (i.e., from 100 to 68,100 words). This random sampling was done with replacement, so each random sample was selected from the total set of all

**Table 1.** Books Included in the Children's Picture-Book Corpus

---

<i>Alexander and the Terrible, Horrible, No Good, Very Bad Day</i> by Judith Viorst	<i>How Do Dinosaurs Say Good Night?</i> by Jane Yolen and Mark Teague
<i>Angelina Ice Skates</i> by Katharine Holabird	<i>How to Train a Train</i> by Jason Carter Eaton
<i>Are You My Mother?</i> by P. D. Eastman	<i>If You Give a Moose a Muffin</i> by Laura Joffe Numeroff
<i>Arnie the Doughnut</i> by Laurie Keller	<i>If You Give a Mouse a Cookie</i> by Laura Joffe Numeroff
<i>Arthur Writes a Story</i> by Marc Brown	<i>I'm a Big Sister</i> by Joanna Cole
<i>A Bad Case of Stripes</i> by David Shannon	<i>The Keeping Quilt</i> by Patricia Polacco
<i>Bark, George</i> by Jules Feiffer	<i>Knuffle Bunny</i> by Mo Willems
<i>Bear Wants More</i> by Karma Wilson	<i>Ladybug Girl at the Beach</i> by David Soman and Jacky Davis
<i>The Berenstain Bears and the Green-Eyed Monster</i> by Stan Berenstain and Jan Berenstain	<i>Lilly's Purple Plastic Purse</i> by Kevin Henkes
<i>The Berenstain Bears Forget Their Manners</i> by Stan Berenstain and Jan Berenstain	<i>Little Blue Truck Leads the Way</i> by Alice Schertle
<i>Blueberries for Sal</i> by Robert McCloskey	<i>The Little Engine That Could</i> by Watty Piper
<i>Bread and Jam for Frances</i> by Russell Hoban	<i>The Little House</i> by Virginia Lee Burton
<i>Brown Bear, Brown Bear, What Do You See?</i> by Bill Martin, Jr.	<i>Llama Llama Home With Mama</i> by Anna Dewdney
<i>Bunny Party</i> by Rosemary Wells	<i>Llama Llama Red Pajama</i> by Anna Dewdney
<i>Caps for Sale</i> by Esphyr Slobodkina	<i>The Lorax</i> by Dr. Seuss
<i>The Carrot Seed</i> by Ruth Krauss	<i>Love You Forever</i> by Sheila McGraw
<i>The Cat in the Hat</i> by Dr. Seuss	<i>Madeline</i> by Ludwig Bemelmans
<i>Charlie and the New Baby</i> by Ree Drummond	<i>Maisy Goes Camping</i> by Lucy Cousins
<i>Chicka Chicka 1-2-3</i> by Bill Martin, Jr., Michael Sampson, and Lois Ehlert	<i>Maisy Goes to the Library</i> by Lucy Cousins
<i>Chicka Chicka Boom Boom</i> by Bill Martin, Jr., and John Archambault	<i>Make Way for Ducklings</i> by Robert McCloskey
<i>Chrysanthemum</i> by Kevin Henkes	<i>Mike Mulligan and His Steam Shovel</i> by Virginia Lee Burton
<i>Click, Clack, Moo: Cows That Type</i> by Doreen Cronin	<i>Miss Rumphius</i> by Barbara Cooney
<i>Clifford at the Circus</i> by Norman Bridwell	<i>The Napping House</i> by Audrey Wood
<i>Cloudy With a Chance of Meatballs</i> by Judi Barrett	<i>No, David!</i> by David Shannon
<i>Corduroy</i> by Don Freeman	<i>Oh, the Places You'll Go</i> by Dr. Seuss
<i>Curious George</i> by H. A. Rey	<i>Olivia</i> by Ian Falconer
<i>Curious George Takes a Job</i> by H. A. Rey	<i>Olivia . . . and the Missing Toy</i> by Ian Falconer
<i>The Day the Crayons Quit</i> by Drew Daywalt	<i>The Other Side</i> by Jacqueline Woodson
<i>Dear Zoo</i> by Rod Campbell	<i>Owl Moon</i> by Jane Yolen
<i>Dinosaur Rescue</i> by Penny Dale	<i>The Paper Bag Princess</i> by Robert N. Munsch
<i>Don't Let the Pigeon Drive the Bus</i> by Mo Willems	<i>Pete the Cat: The Wheels on the Bus</i> by James Dean
<i>Dragons Love Tacos</i> by Adam Rubin	<i>The Pigeon Finds a Hot Dog!</i> by Mo Willems
<i>Duck on a Bike</i> by David Shannon	<i>The Polar Express</i> by Chris Van Allsburg
<i>The Duckling Gets a Cookie!?</i> by Mo Willems	<i>The Runaway Bunny</i> by Margaret Wise Brown
<i>Froggy Goes to Bed</i> by Jonathan London	<i>Show Dog</i> by Meghan McCarthy
<i>The Gardener</i> by Sarah Stewart	<i>A Sick Day for Amos McGee</i> by Philip C. Stead
<i>George and Martha</i> by James Marshall	<i>The Snowy Day</i> by Ezra Jack Keats
<i>The Giving Tree</i> by Shel Silverstein	<i>Stellaluna</i> by Janell Cannon
<i>Goldilicious</i> by Victoria Kann	<i>The Story of Babar</i> by Jean De Brunhoff
<i>Good Night, Gorilla</i> by Peggy Rathman	<i>The Story of Ferdinand</i> by Munro Leaf
<i>Goodnight Moon</i> by Margaret Wise Brown	<i>Sylvester and the Magic Pebble</i> by William Steig
<i>Green Eggs and Ham</i> by Dr. Seuss	<i>The Tale of Peter Rabbit</i> by Beatrix Potter
<i>The Grouchy Ladybug</i> by Eric Carle	<i>That Is Not a Good Idea!</i> by Mo Willems
<i>Guess How Much I Love You</i> by Sam McBratney	<i>There's an Alligator Under My Bed</i> by Mercer Mayer
<i>Harold and the Purple Crayon</i> by Crockett Johnson	<i>This Is Not My Hat</i> by Jon Klassen
<i>Harry the Dirty Dog</i> by Gene Zion	<i>Train</i> by Elisha Cooper
<i>The Hat</i> by Jan Brett	<i>Trashy Town</i> by Andrea Zimmerman and David Clemesha
<i>Horton Hears a Who!</i> by Dr. Seuss	<i>The True Story of the 3 Little Pigs!</i> by Jon Scieszka
	<i>The Very Hungry Caterpillar</i> by Eric Carle
	<i>When Dinosaurs Came With Everything</i> by Elise Broach
	<i>Where the Wild Things Are</i> by Maurice Sendak
	<i>Winter Days in the Big Woods</i> by Laura Ingalls Wilder

---



**Fig. 1.** Results. The graph in (a) shows the mean number of unique words (types) as a function of the total number of words (tokens) in the samples taken from the CHILDES (Child Language Data Exchange System) corpus of child-directed speech and the samples taken from 100 children's picture books. The gray error bars on the CHILDES plot indicate the range of obtained values across the 100 random iterations at each token size. The range of values is not shown on the picture-books plot because the variation is barely visible at this scale. The graph in (b) shows the type and token counts of the 100 picture books and individual conversations in CHILDES, along with the corresponding means from (a).

words in the corpus. This procedure was repeated 100 times for each sample size (i.e., we obtained 100 different random samples from the total corpus at each selection size). The number of unique word types was then counted for each sample.

**Child-directed conversations.** The CHILDES corpus is much larger than our corpus of picture books, so we first needed to obtain an appropriately matched sample of child-directed conversational speech. To accomplish this, we randomly selected contiguous segments from the entire 6.5-million-word corpus, with each contiguous segment matched to the length of 1 of the 100 books in our picture-book corpus. We sampled contiguously because the words in a given book are not independent of each other, just as the words in a conversation are not. If each picture book is considered to have a single topic, then picture books are appropriately compared with conversational child-directed speech in similarly sized segments of single-topic conversations. Our selection method for the CHILDES conversation corpus was thus designed to yield a set of samples comparable to the picture-book corpus—equally sized samples of topically related words.

This selection method yielded a CHILDES sample that consisted of the same number of words as our picture-book corpus. We then applied the same sampling procedure we used with our picture books. We repeated this sampling procedure 100 times (each time with a new, random, length-matched sample from the entirety of CHILDES) and at each sample size counted the number of unique word types in each sample.

## Results

Figure 1a shows the mean number of unique words (types) as a function of the total number of words (tokens) in the samples from the picture books and CHILDES. Note that there was substantially more variability in the CHILDES samples, which was entirely expected, because although those samples were matched on length to the picture-book samples, they were selected from about 6.5 million words. Despite this increased variability, the word-type counts in the CHILDES samples were substantially lower than and almost completely nonoverlapping with the counts in the picture-book samples. When we compared type counts in paired book and text samples at a token size of 100 (e.g., the first sample of 100 book tokens and the first sample of 100 speech tokens, the second sample of 100 book tokens and the second sample of 100 speech tokens), in only 9 of the 100 pairs of random samples of speech and text was the number of unique types higher in speech than in text; when we compared type counts at a token size of 200, in only 2 of the 100 pairs of samples of speech and text were there more unique word types in speech than in text. In all other comparisons of paired samples (i.e., at

token sizes  $\geq 300$ ), there were more unique word types in the samples drawn from text; moreover, for all sample sizes of 300 tokens or greater, the ranges of unique type counts in samples of speech and books were completely nonoverlapping. These estimations of type counts in picture books and child-directed speech clearly show that picture books contain more unique words at a given sample size than does child-directed speech.

In the total sample, picture books contained 1.72 times more unique words than did child-directed speech. It is important to note that the slopes of the lines in Figure 1a are dependent on the size of the corpora in the analysis; thus, the slopes cannot be used to extrapolate the total number of word types a child might hear in a day or year, nor would the ratio of 1.72 remain constant with a different-sized sample of books: When we limited our sample to the first 50 books in our corpus, the ratio between the total number of word types in books and in speech was 1.58, and when we limited our sample to the first 75 books, the ratio was 1.68. These results suggest that the ratio would increase with a larger sample of books. What Figure 1a does show is that the words in 100 picture books and the words in a matched sample of child-directed conversational speech come from different distributions, and the words in picture books are drawn from a more diverse set of vocabulary items compared with the words in child-directed speech. The implications for language learning are clear: Infants whose parents regularly read picture books to them receive more diverse language-learning data—a broader sample of the words in the to-be-learned language—than is indicated by the content of child-directed conversational speech alone.

Individual children participate in individual conversations about individual topics, and parents read children individual books. Thus, a relevant question is how the number of unique word types at various token sizes varies across individual books and individual conversations. That is, perhaps individual books present no greater diversity than individual conversations, but the aggregate diversity across many different books is greater than the aggregate diversity across many different conversations. To better understand how type and token counts in the individual books and conversations in our samples contributed to the observed overall differences in type-token ratios, in Figure 1b we plotted the type and token counts of the 100 books in our corpus and of the 4,432 individual conversations in the CHILDES corpus (each unique CHILDES file containing a single contiguous recording) alongside the means calculated from the randomly sampled speech and text for the corresponding token sizes. The means in Figure 1b are taken from Figure 1a, but the scale is zoomed in, so that the type and token counts of the individual books and the individual conversations can be seen.

Not unexpectedly, the figure shows that both individual books and individual conversations tended to have

fewer types than the means calculated from samples of the same total length. This was expected because, pragmatically, a given book or conversation—if it is a coherent narrative—will be repetitive, so that it will contain fewer unique words than a random selection of a similar number of words. The figure also shows that individual books typically (but of course not always) contained more unique words than did a conversation of the same sample size. This overall pattern, albeit underestimated by the relatively small sample of books, is consistent with two conclusions: First, everyday conversations between parents and young language learners are likely to be more similar to each other than are the individual books that parents might read to children. Second, the words within a single picture book are (typically) more diverse than those within a single conversation.

To better ground these results in children's experiences, consider that conversations are generally limited to here-and-now content, which limits the range of potential topics of conversation. Further, a conversation within an everyday context—for example, mealtime—is likely to have repetitive components day in and day out. Unlike conversations, books are not limited by here-and-now constraints; each book may be different from others in topic or content, opening new domains for discovery and bringing new words into play. Our analyses suggest that although individual books often have more diverse words than do individual conversations, the primary reason that book reading to infants results in a greater diversity of words in language input appears to be that different books sample the words in the language more broadly than do different conversations. Thus, shared book reading, which often begins in infancy, creates a learning environment in which infants and children are exposed to words that they would never have encountered via speech alone. By providing different word distributions than everyday child-directed conversations, shared book reading may play an important role in early language acquisition.

## Discussion

Everyday speech, perhaps because it is constrained by the here-and-now context (Snow & Ninio, 1986) and by the memorial processes that select the words that are produced (Hayes, 1988; see Dell, 1986, and Griffin & Ferreira, 2006, for discussion of the processes that underlie spoken language production), has been repeatedly shown to be more limited in its sampling of language than written prose. The present study shows that the diversity difference between conversation and text also applies to child-directed speech and the child-directed language in picture books. When parents read picture books to infants, the books both bring the exotic into the here and now and, via the text, support the production of a set of relatively uncommon words. The distributional differences between picture books and child-directed speech documented by the present analyses suggest that shared book

reading creates an environment in which children are exposed to more unique words than they would be through speech alone. This finding not only is informative about the data set on which early word learning depends, but may provide important insights into individual differences in early language learning, as well as early differences that are known to be related to later language processing, literacy, and school performance more generally (Bus, van IJzendoorn, & Pellegrini, 1995; Deckner et al., 2006; Scarborough, Dobrich, & Hager, 1991; Sénéchal & LeFevre, 2002).

The extant evidence indicates that reading books to infants is common across large segments of the parent population, but is by no means universal (Bradley, Corwyn, McAdoo, & García Coll, 2001; Raikes et al., 2006; Yarosz & Barnett, 2001; Young et al., 1998). Further, the likelihood of reading to infants and preschool children varies systematically with socioeconomic status (SES; Bradley et al., 2001). Therefore, variability in the frequency with which caregivers read to young children may be an important source of individual differences in language ability. It is well established that the variability in the amount of spoken language that children hear contributes to individual differences in language abilities, with lexical diversity in the input strongly linked to more rapid early vocabulary growth and persisting benefits to vocabulary size (Hart & Risley, 1995; Hoff, 2003; Hoff & Naigles, 2002; Hoff-Ginsberg, 1991; Huttenlocher et al., 2010; Pan et al., 2005; Rowe, 2012; Weisleder & Fernald, 2014; Weizman & Snow, 2001). Hence, parents' speech that is generated in the context of shared book reading may be a key factor contributing to observed individual differences in language ability among young children, and shared book reading may have potential as an intervention for increasing lexical diversity in the learning environments of children (Sharif, Rieber, & Ozuah, 2002).

Given the individual differences in the prevalence of shared book reading across different groups of caregivers (Bradley et al., 2001), the text from books may be a significant factor that contributes to the well-documented differences in early language experiences. Young et al. (1998) and Raikes et al. (2006) both found that the modal tendency (about 50%) for caregivers of 12- to 36-month-old infants is to read to their children at least daily, but some caregivers reported reading to their children multiple times per day, and others never read to their children at all. The mean length of the books in our sample was 680 words—which we round down to 600, to be conservative. At the rate of one book a day, a child would hear more than 219,000 words of text in a year. At the rate of two books a day (Deckner et al., 2006), the child would hear more than 438,000 words of text in a year. The average child recorded by Hart and Risley (1995) heard about 7.3 million words of speech a year. If this typical child were read to once a day, about 3% of his or her linguistic input would be from the text of picture books. If the child were read to twice a day, the estimate would be about 6%. The percentage of speech

derived from text may be even higher if one considers only specifically child-directed speech, which recent work suggests is a better predictor of language learning than all ambient speech (Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2014). Estimates of child-directed speech from Shneidman et al. are in the range of the estimates of all speech reported by Hart and Risley, but Weisleder and Fernald estimated that children from low-SES families hear about 6,000 to 7,000 words of child-directed speech a day; in that case, a single book would constitute nearly 10% of their linguistic input on a given day. Of course, the advantage of adding picture books to the language-learning environment is not necessarily in adding more words per se, but in increasing the diversity of the words in the environment.

The language experience that derives from the text of picture books may be particularly important in light of the linguistic properties of this text and the social context in which it appears. Because of the higher number of unique word tokens (relative to child-directed speech), the contextual diversity of the picture-book text is higher, and such diversity is associated with better learning (Hills et al., 2010; Jones et al., 2012). Thus, the text of picture books may offer particularly good language-learning data. Further, shared book reading is characterized by frequent instances of joint attention (Fletcher et al., 2008; Ninio & Bruner, 1978), and joint attention is known to provide a particularly good language-learning context for children (Farrant & Zubrick, 2012; Tomasello & Farrar, 1986). Despite being a small percentage of the total input, the text of picture books may be disproportionately important for children's language development.

One aspect of shared book reading that we have not addressed is the repetition of books. Caregivers often read the same books to children many times, and the repetition of books would decrease the overall type-token ratio of the language input. Of course, child-directed speech is repetitive too, with similar words used across multiple instances of, for example, mealtime or dressing. With our data, we could not compute the true type-token ratios that a child might encounter in the text of picture books and in child-directed speech. What we have definitively shown, however, is that the text of children's picture books contains more unique word tokens than a properly matched sample of child-directed speech.

The present results directly link early book reading to language acquisition and suggest a possible mechanism by which book reading benefits language acquisition. A very large literature shows that frequent shared book reading in the home is associated with increased vocabulary for children (Farrant & Zubrick, 2012; Fletcher et al., 2008; Karrass & Braungart-Rieker, 2005; Payne et al., 1994; Sénéchal & LeFevre, 2002; Sharif et al., 2002), and with greater success in learning to read and later literacy (Bus et al., 1995; Deckner et al., 2006; Scarborough et al., 1991). There are a number of hypotheses for why reading to children may be associated with these benefits.

Some hypotheses emphasize the pleasant and comforting social aspects of shared book reading (Baker, Scher, & Mackler, 1997). Other hypotheses emphasize that book reading generates extratext caregiver-child dialogue (Deckner et al., 2006; Fletcher et al., 2008; Mol et al., 2008; Ninio & Bruner, 1978; Whitehurst et al., 1988), that books are not limited to discussion of concrete and present items and events (Snow & Ninio, 1986), and that books have a narrative structure (Sulzby, 1985). All these factors likely contribute to enhanced language outcomes. The documented lexical diversity offered by books provides a direct and testable path—though language learning itself. Our results do not change the bottom line for parents: Read to your infants and children. For theorists of language acquisition, our message is that speech derived from text may beneficially expand the data set on which language learning depends.

### Author Contributions

J. L. Montag developed the study concept and design. Data collection and analysis were performed by J. L. Montag under the supervision of M. N. Jones and L. B. Smith. J. L. Montag drafted the manuscript, and M. N. Jones and L. B. Smith provided critical revisions. All authors approved the final version of the manuscript for submission.

### Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

### Funding

This work was supported by National Institutes of Health Grant T32 HD-07475.

### Open Practices

Interested researchers can contact the first author for the picture-book corpus. The complete Open Practices Disclosure for this article can be found at <http://pss.sagepub.com/content/by/supplemental-data>.

### References

- Baker, L., Scher, D., & Mackler, K. (1997). Home and family influences on motivations for reading. *Educational Psychologist, 32*, 69–82.
- Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*, 1844–1867.
- Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*, 1–21.
- Deckner, D. F., Adamson, L. B., & Bakeman, R. (2006). Child and maternal contributions to shared reading: Effects on language and literacy development. *Journal of Applied Developmental Psychology, 27*, 31–41.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

- Dickinson, D. K., Griffith, J. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2012). How reading books fosters language development around the world. *Child Development Research*, 2012, Article 602807. Retrieved from <http://www.hindawi.com/journals/cdr/2012/602807/>
- Farrant, B. M., & Zubrick, S. R. (2012). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language*, 32, 343–364.
- Fletcher, K. L., Cross, J. R., Tanney, A. L., Schneider, M., & Finch, W. H. (2008). Predicting language development in children at risk: The effects of quality and frequency of caregiver reading. *Early Education and Development*, 19, 89–111.
- Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 21–59). London, England: Elsevier.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore, MD: Brookes.
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language*, 27, 572–585.
- Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case of ‘motherese’? *Journal of Child Language*, 15, 395–410.
- Hess, C. W., Sefton, K. M., & Landry, R. G. (1986). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech, Language, and Hearing Research*, 29, 129–134.
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63, 259–273.
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74, 1368–1378.
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73, 418–433.
- Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child Development*, 62, 782–796.
- Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children’s language growth. *Cognitive Psychology*, 61, 343–365.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66, 121–132.
- Karrass, J., & Braungart-Rieker, J. M. (2005). Effects of shared parent–infant book reading on early language acquisition. *Journal of Applied Developmental Psychology*, 26, 133–148.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.
- Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, 19, 7–26.
- Ninio, A. (2011). *Syntactic development, its input and output*. Oxford, England: Oxford University Press.
- Ninio, A., & Bruner, J. (1978). The achievement and antecedents of labelling. *Journal of Child Language*, 5, 1–15.
- Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, 76, 763–782.
- Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, 9, 427–440.
- Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., . . . Rodriguez, E. T. (2006). Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development*, 77, 924–953.
- Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language*, 14, 201–209.
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83, 1762–1774.
- Scarborough, H. S., Dobrich, W., & Hager, M. (1991). Preschool literacy experience and later reading achievement. *Journal of Learning Disabilities*, 24, 508–511.
- Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children’s reading skill: A five-year longitudinal study. *Child Development*, 73, 445–460.
- Sharif, I., Rieber, S., & Ozuah, P. O. (2002). Exposure to Reach Out and Read and vocabulary outcomes in inner city preschoolers. *Journal of the National Medical Association*, 94, 171–177.
- Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, 40, 672–686.
- Snow, C. E., & Ninio, A. (1986). The contracts of literacy: What children learn from learning to read books. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. 116–138). Norwood, NJ: Ablex.
- Sulzby, E. (1985). Children’s emergent reading of favorite storybooks: A developmental study. *Reading Research Quarterly*, 20, 458–481.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57, 1454–1463.
- Weisleder, A., & Fernald, A. (2014). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24, 2143–2152.
- Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children’s vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, 37, 265–279.
- Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture book reading. *Developmental Psychology*, 24, 552–559.
- Willits, J. A., & Jones, M. N. (2015). *What can children learn from 6 million words?* Manuscript submitted for publication.
- Yarosz, D. J., & Barnett, W. S. (2001). Who reads to young children?: Identifying predictors of family reading activities. *Reading Psychology*, 22, 67–81.
- Young, K. T., Davis, K., Schoen, C., & Parker, S. (1998). Listening to parents: A national survey of parents with young children. *Archives of Pediatrics & Adolescent Medicine*, 152, 255–262.