HHS Public Access

Author manuscript

Psychol Sci. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:

Psychol Sci. 2015 September ; 26(9): 1489–1496. doi:10.1177/0956797615594361.

The words children hear: Picture books and the statistics for language learning

Jessica L. Montag[1], **Michael N. Jones**[1], and **Linda B. Smith**[1]

[1]Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana

Abstract

Young children learn language from the speech they hear. Previous work suggests that the 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 to 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.

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 devices are encountered in diverse contexts and when there is more variety in general in the input data (Jones, Johns & Recchia, 2012; Hills et al., 2010). Studies of individual differences in child-directed language also point to diversity in parent talk as a critical predictor of language outcome (Hart & Risley, 1995; Hoff & Naigles, 2002; Hoff, 2003; Huttenlocker, 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

Please direct all correspondence to: Jessica L. Montag, Department of Psychological and Brain Sciences, Indiana University, 1101 E. 10th Street, Bloomington IN, 47405, jlmontag@indiana.edu. Phone: 812-855-6204.

children (Farrant & Zubrick, 2012; Payne, Whitehurst & Angell, 1994; Sénéchal & LaFevre, 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-Reiker, 2005; Young, Davis, Schoen & Parker, 1998). Large representative surveys of parents, indicate that over 50% of parents of infants aged 0 to 5 months report reading books to their infant at least once a week and over 25% of parents of infants 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 child and also 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., 1998). Thus, the words in child-directed books are part of child-directed parent talk.

Analyses of the words present in adult-directed conversations and written texts indicate marked differences in vocabulary choice. 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) while the samples of written texts consisted of 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), suggesting 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 mirrors these differences in lexical diversity, then the language in books may play a significant and 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, then individual differences in shared book reading may play a substantial role in the well-documented individual differences in early language learning and their far-reaching consequences for later language processing (Sénéchal & LaFevre, 2002).

Accordingly, we ask: In terms of the language learning data itself, what do early picture books provide that everyday conversations may 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). 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, yielding four unique words. Type-token counts and 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).

Methods

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 consists 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, the titles were selected from lists of librarian-recommended picture books, amazon.com bestsellers, 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 0 to 60 months, and thus within the age range of the conversational CHILDES 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 aged 0–60 months to match the age range to the intended age range of the picture books. Our sample was comprised of 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 (1) remove a number of the special transcription characters and other artifacts of the CHILDES coding system and (2) 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 by a single sample or measure (Sefton & Landry, 1986; Richards, 1987). Accordingly, our principle dependent measure is number of unique word types across multiple sample (token) sizes for child-directed conversational speech and child-directed text. To obtain a distribution of type counts as a function of tokens, we randomly sampled sets of words of various sizes from the two corpora.

Picture Books—The different sized random samples were constructed by taking progressively larger random samples that increased in increments of 100 words. This random sampling was done with replacement, so each random sample was selected from the total set of all words in the corpus. This procedure was repeated 100 times for each sample size, yielding 100 different random samples from the total corpus at each selection size. The number of unique word types was then counted for each sample at each selection size.

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 of CHILDES from the entire 6.5 million word corpus with each contiguous segment matched to the length of each 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 could be considered to consist of a single topic then the appropriate comparison conversational child-directed speech is to compare 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 sample of CHILDES that consisted of the same number of words as our picture books corpus. We then applied the same sampling procedure we used with our picture books to calculate the number of unique words in child-directed speech at increasing sample sizes. We repeated this sampling procedure 100 times (each time with a new, random, length-matched sample from the entirety of CHILDES) and computed the number of unique word types at each selection size.

Results

Figure 1a below shows the mean number of types (unique words) at each token sample size from both the picture books and CHILDES. The error bars on the CHILDES plot refer to the *range* of obtained values across the 100 random iterations through the corpora. No error bars are plotted on the Picture Books plot because they are barely visible at this scale. This is entirely expected, because the picture book length-matched samples from CHILDES were selected from about 6.5 million words, so there was substantially more variability in our CHILDES samples. Despite this increased variability in the CHILDES sample, the word type counts of CHILDES were substantially lower than and almost completely non-overlapping with that of the picture books. When comparing type counts in paired book and text samples (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), at a token size of 100, in only nine of the 100 random samples of speech and text was the number of unique types on speech higher than the number in text and at a token size of 200, in only two of the 100 samples of speech and text were there more unique word types in speech. In all other comparison pairs, not only there were more unique word types in the samples drawn from text, but in all samples of 300 tokens or greater, the ranges of unique type counts in samples of speech and books were completely non-overlapping. 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 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 are dependent upon the sample size from which the samples were drawn, thus they 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: If we limit our sample to the first 50 books in the sample, the ratio between word tokens and types is 1.58 and limited to the first 75 books, the ratio is 1.68, suggesting that the ratio would in fact *increase* with a larger sample of books. What this figure 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 distribution from which the words in picture books are drawn contains

a more diverse set of vocabulary items relative to child-directed speech. The implications for language learning are clear: The language learning data for infants who are regularly read picture books by their parents is different – more diverse, a broader sample of the words in the to-be-learned language –than would be indicated from 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 conservations but the aggregate diversity across parental reading of 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 contribute to the observed overall differences in type-token ratios, we plotted the overall type-token counts calculated from the randomly sampled speech and text, alongside the type and token counts of the 100 books that comprised the corpus as well as each of the 4,568 individual conversations that comprise CHILDES (each unique CHILDES file that contains a single contiguous recording). These values are plotted in Figure 1b.

The large black and gray points refer to the mean type and token counts in Figure 1a, but in Figure 1b the scale is zoomed in, enabling the visualization of the type and token counts of the individual books and the individual conversations. First, and not unexpectedly, both individual books and conversations tend to have fewer types than the mean samples of the same total length. This is expected because pragmatically, books and conversations –if they are coherent narratives –will be individually repetitive, so a real conversation of a given length will contain fewer unique words than a random selection of a similar number of words. Second, individual books typically (but of course not always) contain more unique words given the sample size than does a conversation (the small black dots indicating the number of types in individual books are mostly above small gray dots indicating the types in conversations). This overall pattern, albeit underestimated by the present relatively small sample of books, is consistent with two conclusions: (1) everyday conversations between parents and young language learners are likely be more similar to each other than are the individual books that parents might read to children and (2) 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 more different than the others in topic or content, with each new book opening new domains for discovery and new words. These 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 the input appears to be because different books sample the words in the language more broadly than do different 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 statistics 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; 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 brings the exotic into the here-and-now and, via the text, support the production of a set of less common 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 informs the data set on which early word learning depends, but may provide important insights into individual differences in early language learning, and early differences that are known to be related to later language processing, literacy, and school performance more generally (Bus, Ijzendoorn & Pellegrini, 1995; Deckner et al., 2006; Scarborough, Dobrich & Hager, 1991; Sénéchal & LaFevre, 2002).

The extant evidence indicates that book reading to infants is common across large segments of the parent population, but is by no means universal (Bradley, et al., 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 socio-economic status (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 a child hears contributes to individual differences in language abilities, with lexical diversity strongly linked to more rapid vocabulary growth and to later language acquisition (Hart & Risley, 1995; Hoff, 2003; Hoff & Naigles, 2002; Hoff-Ginsberg, 1991; Huttenlocker, et al., 2010; Pan, et al., 2005; Rowe, 2012; Weizman & Snow, 2001; Weisleder & Fernald, 2014). Hence, parent speech that is generated in the context of shared book reading may be a key factor of observed individual differences in language ability among young children and also a potential intervention for increasing lexical diversity in the learning environments of children (Sharif, Reiber & 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 in the linguistic experiences that create the well-documented differences in early language experiences. Young et al. (1998) and Raikes et al. (2006) both find 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 report reading as many as multiple times per day while others

never read to their children at all. The mean book length in our sample was 680 words—which we round down to 600, to be conservative. At a rate of one book a day, a child would hear over 218,000 words of text in a year. At the rate of two books a day (Deckner, et al., 2006), the child would hear over 436,000 words of text a year. The average child recorded by Hart and Risley (1995) heard about 7.3 million words of speech a year, so for a typical child read to once a day, about 3% of their linguistic input would be from the text of picture books. For a child read to twice a day, the estimate is about 6%. These estimates of speech from text may be even higher when considering only 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 consistent with the estimates we report above, but if, as Weisleder and Fernald estimate, children from low-SES families hear about 6,000–7,000 words of child-directed speech a day, a single book would comprise nearly 10% of their linguistic input. Of course the advantage of adding picture books to the language learning environment is not necessarily in more words per se, but in the increased diversity of those words relative to conversational speech.

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 this text appears. Due to the higher number of unique word tokens, the contextual diversity of the picture book text is higher, which is associated with better learning (Hills et al., 2010; Jones et al., 2012) so 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), which are known to be particularly good language learning contexts 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 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 cannot compute the true type-token ratios in text of picture books or child-directed speech that a child might encounter. What we do definitively show 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 links the benefits of early book reading to language acquisition. A very large literature shows that homes in which shared book reading is common are associated with increased vocabulary for children (Farrant & Zubrick, 2012; Fletcher et al., 2008; Karrass & Braugart-Rieker, 2005; Sénéchal & LaFevre, 2002; Payne, et al., 1994; Sharif, et al., 2002), and with greater success in learning to read and later literacy (Bus, Ijzendoorn & Pellegrini, 1995; Deckner et al., 2006; Scarborough, Dobrich & Hager, 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 the extra-text

caregiver-child dialogue that book reading generates (Deckner, et al., 2006; Fletcher et al., 2008; Mol, et al., 2008; Ninio & Bruner, 1978; Whitehurst et al., 1998), that books are not limited to discussion of concrete and present items and events (Snow & Ninio, 1986) and the narrative structure of books (Sulzby, 1985). All these factors likely contribute. The documented differences in lexical diversity provide a direct and testable path –though language learning itself. The bottom line from a parent's perspective is the same –read to your infants and children. For theorists of language acquisition, speech derived from text may beneficially expand the data set on which language learning depends.

Acknowledgments

This work was supported by NIH Grant T32 HD-07475.

References

Baker L, Scher D, Mackler K. Home and family influences on motivations for reading. Educational Psychologist. 1997; 32:69–82.

Bradley RH, Corwyn RF, McAdoo HP, García Coll C. The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. Child Development. 2001; 72:1844–1867. [PubMed: 11768149]

Chafe W, Tannen D. The relationship between written and spoken language. Annual Review of Anthropology. 1987; 16:383–407.

Deckner DF, Adamson LB, Bakeman R. Child and maternal contributions to shared reading: Effects on language and literacy development. Journal of Applied Developmental Psychology. 2006; 27(1): 31–41.

Dell GS. A spreading-activation theory of retrieval in sentence production. Psychological Review. 1986; 93:283. [PubMed: 3749399]

DeVito JA. Psychogrammatical factors in oral and written discourse by skilled communicators. Speech Monographs. 1966; 33:73–76.

Dickinson DK, Griffith JA, Golinkoff RM, Hirsh-Pasek K. How reading books fosters language development around the world. Child Development Research. 2012; 2012:1–15.

Farrant BM, Zubrick SR. Early vocabulary development: The importance of joint attention and parent-child book reading. First Language. 2012; 32:343–364.

Griffin, ZM.; Ferreira, VS. Properties of spoken language production. In: Traxler, MJ.; Gernsbacher, MA., editors. Handbook of psycholinguistics. 2. London: Elsevier; 2006. p. 21-59.

Fletcher KL, Cross JR, Tanney AL, Schneider M, Finch WH. Predicting language development in children at risk: The effects of quality and frequency of caregiver reading. Early Education and Development. 2008; 19:89–111.

Hart, B.; Risley, TR. Meaningful differences in the everyday experiences of young American children. Baltimore: Brookes; 1995.

Hayes DP. Speaking and writing: Distinct patterns of word choice. Journal of Memory and Language. 1988; 27:572–585.

Hayes DP, Ahrens MG. Vocabulary simplification for children: A special case of 'motherese'? Journal of Child Language. 1988; 15:395–410. [PubMed: 3209647]

Hess CW, Sefton KM, Landry RG. Sample size and type-token ratios for oral language of preschool children. Journal of Speech, Language, and Hearing Research. 1986; 29:129–134.

Hills TT, Maouene J, Riordan B, Smith LB. The associative structure of language: Contextual diversity in early word learning. Journal of Memory and Language. 2010; 63:259–273. [PubMed: 20835374]

Hoff E. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development. 2003; 74:1368–1378. [PubMed: 14552403]

Hoff E, Naigles L. How children use input to acquire a lexicon. Child Development. 2002; 73:418–433. [PubMed: 11949900]

Huttenlocher J, Waterfall H, Vasilyeva M, Vevea J, Hedges LV. Sources of variability in children's language growth. Cognitive psychology. 2010; 61:343–365. [PubMed: 20832781]

Jones MN, Johns BT, Recchia G. The role of semantic diversity in lexical organization. Canadian Journal of Experimental Psychology. 2012; 66:121–132.

Karrass J, Braungart-Rieker JM. Effects of shared parent–infant book reading on early language acquisition. Journal of Applied Developmental Psychology. 2005; 26:133–148.

MacWhinney, B. The CHILDES Project: Tools for analyzing talk. 3. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

Mol SE, Bus AG, de Jong MT, Smeets DJ. Added value of dialogic parent–child book readings: A meta-analysis. Early Education and Development. 2008; 19:7–26.

Montag JL, Smith LB, Jones MN. Considerations for the use of word type and token counts and ratios in language acquisition research. in preparation.

Ninio, A. Syntactic development, its input and output. Oxford University Press; 2011.

Ninio A, Bruner J. The achievement and antecedents of labelling. Journal of Child Language. 1978; 5:1–15.

Pan BA, Rowe ML, Singer JD, Snow CE. Maternal correlates of growth in toddler vocabulary production in low-income families. Child development. 2005; 76:763–782. [PubMed: 16026495]

Payne AC, Whitehurst GJ, Angell AL. The role of home literacy environment in the development of language ability in preschool children from low-income families. Early Childhood Research Quarterly. 1994; 9:427–440.

Raikes H, Alexander Pan B, Luze G, Tamis-LeMonda CS, Brooks-Gunn J, Constantine J, Rodriguez ET. Mother–child bookreading in low-income families: Correlates and outcomes during the first three years of life. Child Development. 2006; 77:924–953. [PubMed: 16942498]

Richards B. Type/token ratios: what do they really tell us? Journal of Child Language. 1987; 14:201–209. [PubMed: 3611238]

Rowe ML. A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development. Child Development. 2012; 83:1762–1774. [PubMed: 22716950]

Shneidman LA, Arroyo ME, Levine SC, Goldin-Meadow S. What counts as effective input for word learning? Journal of Child Language. 2013; 40:672–686. [PubMed: 22575125]

Sénéchal M, LeFevre JA. Parental involvement in the development of children's reading skill: A five-year longitudinal study. Child Development. 2002; 73:445–460. [PubMed: 11949902]

Sharif I, Rieber S, Ozuah PO. Exposure to Reach Out and Read and vocabulary outcomes in inner city preschoolers. Journal of the National Medical Association. 2002; 94:171. [PubMed: 11918387]

Snow, CE.; Ninio, A. The contracts of literacy: What children learn from learning to read books. In: Teale, WH.; Sulxby, E., editors. Emergent literacy: Writing and reading. Norwood, NJ: Ablex; 1986. p. 116-138.

Sulzby E. Children's emergent reading of favorite storybooks: A developmental study. Reading Research Quarterly. 1985; 20:458–481.

Tomasello M, Farrar MJ. Joint attention and early language. Child Development. 1986:1454–1463. [PubMed: 3802971]

Weisleder A, Fernald A. Talking to children matters early language experience strengthens processing and builds vocabulary. Psychological Science. 2014; 24:2143–2152. [PubMed: 24022649]

Weizman ZO, Snow CE. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. Developmental Psychology. 2001; 37:265. [PubMed: 11269394]

Whitehurst GJ, Falco FL, Lonigan CJ, Fischel JE, DeBaryshe BD, Valdez-Menchaca MC, Caulfield M. Accelerating language development through picture book reading. Developmental Psychology. 1988; 24:552.

Willits JA, Jones MN. What can children learn from 6 million words?. in review.

Yarosz DJ, Barnett WS. Who reads to young children?: Identifying predictors of family reading activities. Reading Psychology. 2001; 22:67–81.

Young KT, Davis K, Schoen C, Parker S. Listening to parents: a national survey of parents with young children. Archives of Pediatrics & Adolescent Medicine. 1998; 152:255–262. [PubMed: 9529463]

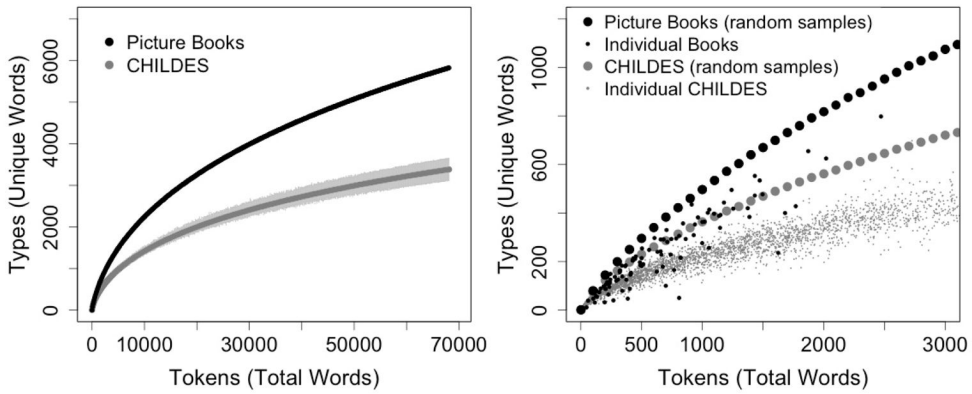

Figure 1.

Figure 1a (left): Mean number of unique words (types) as a function of number of total words (tokens) in child-directed speech and children's picture books.

Figure 1b (right): Large black and gray points are the same type and token counts plotted in Figure 1a. Small black dots refer to the individual type and token counts of the 100 picture books and small gray dots refer to type and token counts of individual conversations in CHILDES.

Table 1

Books included in the children's picture book corpus

A Bad Case of Stripes by David Shannon
A Sick Day for Amos McGee by Philip C. Stead
Alexander and the Terrible, Horrible, No Good, Very Bad Day by Judith Viorst
Angelina Ice Skates by Katharine Holabird
Are You My Mother? by P.D. Eastman
Arnie the Doughnut by Lourie Keller
Arthur Writes a Story by Marc Brown
Bark, George by Jules Feiffer
Bear Wants More by Karma Wilson
Blueberries for Sal by Robert McCloskey
Bread and Jam for Frances by Russell Hoban
Brown Bear, Brown Bear, What Do You See? by Bill Martin Jr
Bunny Party by Rosemary Wells
Caps for Sale by Esphyr Slobodkina
Charlie and the New Baby by Ree Drummond
Chicka Chicka 1-2-3 by Bill Martin Jr., Michael Sampson & Lois Ehlert
Chicka Chicka Boom Boom by Bill Martin Jr. & John Archambault
Chrysanthemum by Kevin Henkes
Click, Clack, Moo Cows that Type by Doreen Cronin
Clifford at the Circus by Norman Bridwell
Cloudy With a Chance of Meatballs by Judi Barrett
Corduroy by Don Freeman
Curious George by H.A. Rey
Curious George Takes a Job by H.A. Ray
Dear Zoo by Rod Campbell
Dinosaur Rescue by Penny Dale
Don't Let the Pigeon Drive the Bus by Mo Willems
Dragons Love Tacos by Adam Rubin
Duck on a Bike by David Shannon
Froggy Goes to Bed by Jonathan London
George and Martha by James Marshall
Goldilicious by Victoria Kann
Good Night Gorilla by Peggy Rathman
Goodnight Moon by Margaret Wise Brown
Green Eggs and Ham by Dr. Seuss
Guess How Much I Love You by Sam McBratney
Harold and the Purple Crayon by Crockett Johnson
Harry the Dirty Dog by Gene Zion
Horton Hears a Who! by Dr. Seuss
How Do Dinosaurs Say Good Night? by Jane Yolen & Mark Teague

How to Train a Train by Jason Carter Eaton

I'm a Big Sister by Joanna Cole

If You Give a Moose a Muffin by Laura Joffe Numeroff

If You Give a Mouse a Cookie by Laura Joffe Numeroff

Knuffle Bunny by Mo Willems

Ladybug Girl at the Beach by David Soman & Jacky Davis

Lilly's Purple Plastic Purse by Kevin Henkes

Little Blue Truck Leads the Way by Alice Schertle

Llama Llama Home with Mama by Anna Dewdney

Llama Llama Red Pajama by Anna Dewdney

Love You Forever by Sheila McGraw

Madeline by Ludwig Bemelmans

Maisy Goes Camping by Lucy Cousins

Maisy Goes to the Library by Lucy Cousins

Make Way for Ducklings by Robert McCloskey

Mike Mulligan and his Steam Shovel by Virginia Lee Burton

Miss Rumphius by Barbara Cooney

No, David! by David Shannon

Oh, the Places You'll Go by Dr. Seuss

Olivia by Ian Falconer

Olivia… and the Missing Toy by Ian Falconer

Own Moon by Jane Yolen

Pete the Cat: The Wheels on the Bus by James Dean

Show Dog by Meghan McCarthy

Stellaluna by Janell Cannon

Sylvester and the Magic Pebble by William Steig

That Is Not a Good Idea! by Mo Willems

The Berenstain Bears and the Green-Eyed Monster by Stan & Jan Berenstain

The Berenstain Bears Forget Their Manners by Stan & Jan Berenstain

The Carrot Seed by Ruth Krauss

The Cat in The Hat by Dr. Seuss

The Day the Crayons Quit by Drew Daywalt

The Duckling Gets a Cookie!? by Mo Willems

The Gardener by Sarah Stewart

The Giving Tree by Shel Silverstein

The Grouchy Ladybug by Eric Carle

The Hat by Jan Brett

The Keeping Quilt by Patricia Polacco

The Little Engine That Could by Watty Piper

The Little House by Virginia Lee Burton

The Lorax by Dr. Seuss

The Napping House by Audrey Wood

The Other Side by Jacqueline Woodson

The Paper Bag Princess by Robert N. Munsch

The Pigeon Finds a Hot Dog! by Mo Willems

The Polar Express by Chris Van Allsburg

The Runaway Bunny by Margaret Wise Brown

The Snowy Day by Ezra Jack Keats

The Story of Babar by Jean De Brunhoff

The Story of Ferdinand by Munro Leaf

The Tale of Peter Rabbit

The True Story of the 3 Little Pigs! by Jon Scieszka

The Very Hungry Caterpillar by Eric Carle

There's an Alligator Under My Bed by Mercer Mayer

This is Not My Hat by Jon Klassen

Train by Elisha Cooper

Trashy Town by Andrea Zimmerman & David Clemesha

When Dinosaurs Came with Everything by Elise Broach

Where the Wild Things Are by Maurice Sendak

Winter Days in the Big Woods by Laura Ingalls Wilder