

# FIRST LANGUAGE

VOLUME 59 - ISSUE 1 - NUMBER 04 - FEBRUARY 2008



## How the characteristics of words in child-directed speech differ from adult-directed speech to influence children's productive vocabularies

Journal:	<i>First Language</i>
Manuscript ID	FLA-22-0024.R3
Manuscript Type:	Original Manuscript
Keywords:	child-directed speech, word learning, motherese, caregiver speech, vocabulary
Abstract:	<p>Child-directed speech has long been known to influence children's vocabulary learning. However, while we know that caregiver utterances differ from those directed at adults in various ways, little is known about any differences in the lexical properties of child-directed and adult-directed utterances. We compare over half a million word tokens from adult speech directed at children (from caregiver-child transcriptions) to the same quantity directed at adults. We show that child-directed speech contains greater numbers of words that are lower in phonemic length, higher in frequency, lower in phonotactic probability, and higher in neighborhood density than adult-directed speech; furthermore, child-directed speech explains over twice the variability of children's productive noun vocabularies than adult-directed speech. These findings indicate that children's word production is clearly influenced by the characteristics of the words spoken directly to them and that researchers need to be wary of using adult-directed language corpora when calculating lexical measures.</p>

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1 Child- and adult-directed speech

1

2  
3 How the characteristics of words in child-directed speech differ from adult-directed speech to  
4  
5 influence children's productive vocabularies  
6  
7  
8  
9  
10

11 Gary Jones<sup>1</sup>

12 Francesco Cabiddu<sup>2</sup>

13  
14  
15  
16  
17 Doug J. K. Barrett<sup>3</sup>

18  
19  
20 Antonio Castro<sup>1</sup>

21  
22  
23 Bethany Lee<sup>1</sup>

24  
25  
26 <sup>1</sup>Nottingham Trent University, UK

27  
28  
29 <sup>2</sup>Cardiff University, UK

30  
31  
32 <sup>3</sup>University of Leicester, UK  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

#### 44 Author Note

45  
46 Correspondence concerning this article should be addressed to Gary Jones, Department of  
47  
48 Psychology, Nottingham Trent University, 50 Shakespeare Street, Nottingham NG1 4FQ  
49  
50 (UK). E-mail: gary.jones@ntu.ac.uk. Telephone: +44 (115) 848 2422.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Child- and adult-directed speech

## Abstract

Child-directed speech has long been known to influence children's vocabulary learning. However, while we know that caregiver utterances differ from those directed at adults in various ways, little is known about any differences in the lexical properties of child-directed and adult-directed utterances. We compare over half a million word tokens from adult speech directed at children (from caregiver-child transcriptions) to the same quantity directed at adults. We show that child-directed speech contains greater numbers of words that are lower in phonemic length, higher in frequency, lower in phonotactic probability, and higher in neighborhood density than adult-directed speech; furthermore, child-directed speech explains over twice the variability of children's productive noun vocabularies than adult-directed speech. These findings indicate that children's word production is clearly influenced by the characteristics of the words spoken directly to them and that researchers need to be wary of using adult-directed language corpora when calculating lexical measures.

Keywords: child-directed speech; word learning; motherese; caregiver speech; vocabulary.

## Introduction

Child-directed speech differs from adult-directed speech in ways that influence children's language acquisition, such as exaggerated intonation and shorter utterances (e.g., Saxton, 2009). The lexical characteristics of child-directed speech also influence child language, with early word production being influenced by factors such as spoken frequency of occurrence and word length (e.g., Jones, Cabiddu, Andrews & Rowland, 2021). **However, it is unclear whether the lexical characteristics of words learned in early childhood are more similar to the lexical characteristics of words in child-directed speech than adult-directed speech, and as a consequence, whether such lexical characteristics facilitate early word learning more so for child-directed speech than adult-directed speech.** In this paper we provide a detailed comparison of the characteristics of words used in the vocabularies of child-directed and adult-directed speech, together with how those characteristics map onto children's productive vocabularies. We will show how children's early expressive vocabularies are influenced by the lexical characteristics of child-directed speech in particular, suggesting that the characteristics of words spoken to children promotes their production of some words over others.

Adults change the way they talk depending on who they are speaking to. For example, when native-speaking adults talk to other adults for whom the native language is foreign, spoken language is simplified (e.g., Rodriguez-Cuadrado, Baus & Costa, 2018). It should therefore come as no surprise that if an adult is speaking to a child, their spoken language will be different to that used when speaking to another adult. This is particularly the case when addressing infants (e.g., Saxton, 2009; Suttora et al., 2017). First, intonation is more pronounced such that communicative intent is clearer (e.g., Fernald, 1989), with child-directed speech being characterized by a slower speech rate and greater variability in pitch (e.g., Thiessen, Hill & Saffran, 2005). These alterations to typical adult speech are believed to

## Child- and adult-directed speech

aid the discriminability of vowels and potentially consonants (Liu, Kuhl & Tsao, 2003) with a reduction in speech rate potentially being a key factor (McMurray, Kovack-Lesh, Goodwin & McEchron, 2013). Second, child-directed speech also involves fewer words than adult-directed speech (Saxton, 2009), with speech length being adapted to suit the linguistic competence of the child (Roy, Frank & Roy, 2009). Third, sentence structure and word content are changed from what would be used when talking to other adults. For example, Furrow, Nelson and Benedict (1979) showed that **speech directed towards 18- to 27-month old children** had greater numbers of *wh*-questions and fewer declaratives than adult-directed speech, while You, Bickel, Daum and Stoll (2021) showed that child-directed speech allowed causal meaning to be inferred based on the co-occurrence of neighboring words, which was not possible with adult-directed speech.

The changes in the spoken language used when adults address children rather than other adults may well influence children's vocabulary acquisition. Weisleder and Fernald (2013) measured both child-directed speech and adult-directed (overheard) speech involving 19-month old infants within 29 families, finding only the former to be linked to infant vocabulary size 6 months later. Changes in the length of spoken utterances seems to have an effect on child productive vocabularies, with shorter utterances directed to 18-month-old infants relating to larger productive vocabularies at 27-months (Furrow, Nelson & Benedict, 1979) while for speech directed to older infants (18- to 29-month old) the relationship reverses (Hoff & Naigles, 2002)<sup>1</sup>. With regard to intonation, 7-month old infants are more able to identify word boundaries in a continuous stream of nonsense words when the words are spoken with exaggerated prosody that is consistent with child-directed speech, rather than that of adult-directed speech (Thiessen, Hill & Saffran, 2005). **Since identifying words in a**

---

<sup>1</sup> This apparent contradiction is likely related to changes in the interaction taking place (Dong, Gu & Vigliocco, 2021; see also Golinkoff, Can, Soderstrom & Hirsh-Pasek, 2015).

1  
2  
3 continuous stream of speech is part of the challenge for word learning, one may expect the  
4 exaggerated prosody of child-directed speech to facilitate vocabulary acquisition. There are  
5 also differences in word use across child-directed and adult-directed speech, and these  
6 differences may also have consequences for children's word learning. For example, Hills  
7 (2013) measured contextual diversity and word repetitions in child-directed speech and adult-  
8 directed speech, finding that child-directed speech was less diverse and more repetitive than  
9 adult-directed speech. Moreover, children's word learning was better predicted by child-  
10 directed speech. For example, Hills  
11 (2013) measured contextual diversity and word repetitions in child-directed speech and adult-  
12 directed speech, finding that child-directed speech was less diverse and more repetitive than  
13 adult-directed speech. Moreover, children's word learning was better predicted by child-  
14 directed speech. Moreover, children's word learning was better predicted by child-  
15 directed speech. Moreover, children's word learning was better predicted by child-  
16 directed speech. Moreover, children's word learning was better predicted by child-  
17 directed speech. Moreover, children's word learning was better predicted by child-  
18 directed speech. Moreover, children's word learning was better predicted by child-  
19 directed speech. Moreover, children's word learning was better predicted by child-  
20 directed speech. Moreover, children's word learning was better predicted by child-  
21 directed speech. Moreover, children's word learning was better predicted by child-  
22 directed speech. Moreover, children's word learning was better predicted by child-  
23 directed speech. Moreover, children's word learning was better predicted by child-  
24 directed speech. Moreover, children's word learning was better predicted by child-  
25 directed speech. Moreover, children's word learning was better predicted by child-  
26 directed speech. Moreover, children's word learning was better predicted by child-  
27 directed speech. Moreover, children's word learning was better predicted by child-  
28 directed speech. Moreover, children's word learning was better predicted by child-  
29 directed speech. Moreover, children's word learning was better predicted by child-  
30 directed speech. Moreover, children's word learning was better predicted by child-  
31 directed speech. Moreover, children's word learning was better predicted by child-  
32 directed speech. Moreover, children's word learning was better predicted by child-  
33 directed speech. Moreover, children's word learning was better predicted by child-  
34 directed speech. Moreover, children's word learning was better predicted by child-  
35 directed speech. Moreover, children's word learning was better predicted by child-  
36 directed speech. Moreover, children's word learning was better predicted by child-  
37 directed speech. Moreover, children's word learning was better predicted by child-  
38 directed speech. Moreover, children's word learning was better predicted by child-  
39 directed speech. Moreover, children's word learning was better predicted by child-  
40 directed speech. Moreover, children's word learning was better predicted by child-  
41 directed speech. Moreover, children's word learning was better predicted by child-  
42 directed speech. Moreover, children's word learning was better predicted by child-  
43 directed speech. Moreover, children's word learning was better predicted by child-  
44 directed speech. Moreover, children's word learning was better predicted by child-  
45 directed speech. Moreover, children's word learning was better predicted by child-  
46 directed speech. Moreover, children's word learning was better predicted by child-  
47 directed speech. Moreover, children's word learning was better predicted by child-  
48 directed speech. Moreover, children's word learning was better predicted by child-  
49 directed speech. Moreover, children's word learning was better predicted by child-  
50 directed speech. Moreover, children's word learning was better predicted by child-  
51 directed speech. Moreover, children's word learning was better predicted by child-  
52 directed speech. Moreover, children's word learning was better predicted by child-  
53 directed speech. Moreover, children's word learning was better predicted by child-  
54 directed speech. Moreover, children's word learning was better predicted by child-  
55 directed speech. Moreover, children's word learning was better predicted by child-  
56 directed speech. Moreover, children's word learning was better predicted by child-  
57 directed speech. Moreover, children's word learning was better predicted by child-  
58 directed speech. Moreover, children's word learning was better predicted by child-  
59 directed speech. Moreover, children's word learning was better predicted by child-  
60 directed speech. Moreover, children's word learning was better predicted by child-

### Current study

27 While there are many examples of how child-directed speech may benefit children's  
28 vocabularies, few if any compare the lexical properties of child-directed and adult-directed  
29 speech. The current study examines four well-established lexical characteristics (word length,  
30 frequency of occurrence, neighborhood density, and phonotactic probability) to see if they  
31 differ across child-directed and adult-directed speech and to what extent any differences are  
32 predictive of children's early vocabularies. The current study examines four well-established lexical characteristics (word length,  
33 frequency of occurrence, neighborhood density, and phonotactic probability) to see if they  
34 differ across child-directed and adult-directed speech and to what extent any differences are  
35 predictive of children's early vocabularies. The current study examines four well-established lexical characteristics (word length,  
36 frequency of occurrence, neighborhood density, and phonotactic probability) to see if they  
37 differ across child-directed and adult-directed speech and to what extent any differences are  
38 predictive of children's early vocabularies. The current study examines four well-established lexical characteristics (word length,  
39 frequency of occurrence, neighborhood density, and phonotactic probability) to see if they  
40 differ across child-directed and adult-directed speech and to what extent any differences are  
41 predictive of children's early vocabularies.

42 A number of metrics have been used to characterize the lexical properties of the early  
43 words in child vocabularies: word length<sup>2</sup>, word frequency, neighborhood density, and  
44 phonotactic probability. Not surprisingly, early vocabularies contain more short than long  
45 words (e.g., Maekawa & Storkel, 2006; Storkel, 2009). This is consistent with a wealth of  
46 memory and related literature where both children and adults have greater difficulty recalling  
47 long words and long nonsense words (e.g., Baddeley, Thomson, & Buchanan, 1975;  
48 Gathercole & Baddeley, 1989). For example, numerous studies show that children are more  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58

---

59 <sup>2</sup> Phonemic or syllabic word length (i.e., we do not consider spoken duration).  
60

## Child- and adult-directed speech

6

1  
2  
3 accurate at repeating nonsense words when the nonsense words are short rather than long  
4  
5 (e.g., Jones, Tamburelli, Watson, Gobet, & Pine, 2010; Jones & Witherstone, 2011).

7  
8 However, we are unaware of whether child-directed speech contains a proportionally larger  
9  
10 number of short than long words relative to adult-directed speech. It seems plausible that this  
11  
12 would be the case on the basis of mean length of utterance being smaller in child-directed  
13  
14 speech than adult-directed speech (e.g., Bernstein Ratner & Rooney, 2001), but as far as we  
15  
16 are aware, no study has compared word length between the two types of speech.  
17  
18

19  
20 Child vocabularies are also dominated by words that occur frequently in the native  
21  
22 language. For example, Huttenlocher et al. (1991) showed that as the frequency by which  
23  
24 particular content words occurred increased in parental speech, the age at which their children  
25  
26 produced those same words decreased. A similar pattern is found for verbs (Naigles & Hoff-  
27  
28 Ginsberg, 1998) and more generally across word categories. Goodman, Li and Dale (2008)  
29  
30 examined common nouns, people words, verbs, adjectives, closed-class and other words  
31  
32 using 3.8 million tokens from mother-child transcripts on CHILDES (MacWhinney, 2000),  
33  
34 showing that higher parental frequency of words in each category led to earlier child  
35  
36 production of those words. Similarly, Swingley and Humphrey (2018) showed that words  
37  
38 occurring with a high rather than low frequency in the maternal input were far more likely to  
39  
40 be present in the child's productive vocabulary. In the work above, word frequency is  
41  
42 calculated on the basis of child-directed speech. In other work, word frequency also  
43  
44 influences vocabulary acquisition when it is measured based on adult-directed written  
45  
46 language (e.g., Stokes, 2010). That said, it is not clear whether child-directed speech contains  
47  
48 greater numbers of high frequency words over adult-directed language. As far as we are  
49  
50 aware, the only work to compare the two was Hayes and Ahrens (1988) who found no word  
51  
52 frequency differences between speech directed at young children (up to two years of age) and  
53  
54  
55  
56  
57  
58  
59  
60



## Child- and adult-directed speech

7

1  
2  
3 speech aimed at older children (two to twelve years). However, their analysis used a  
4  
5 relatively small number of speech tokens and needs to be verified using larger samples.  
6  
7

8  
9  
10 Child vocabularies consist of proportionally more dense neighborhood words than  
11  
12 adult vocabularies (Coady & Aslin, 2003). For any given word, a neighbor is another word  
13  
14 that differs by the addition, deletion or substitution of one sound e.g., *slit*, *it*, and *lot* are  
15  
16 neighbors of *lit*. Based on examination of MacArthur Child Development Inventories  
17  
18 (MCDIs, checklists that allow caregivers to indicate words known and words produced by  
19  
20 infants), Storkel (2009) showed that children aged 16- to 28-months knew a greater number  
21  
22 of nouns with many neighbors than they did those with few neighbors. The effect of  
23  
24 neighborhood density appears to be strong. For example, Stokes (2010) examined all  
25  
26 monosyllabic words in the MCDI and found 47% of the variance in 24-30 month old  
27  
28 children's vocabularies was accounted for by their neighborhood density in adult language.  
29  
30 Knowledge of similar sounding words scaffold the learning of a novel word (see Jones et al.,  
31  
32 2021); if caregiver speech contains a greater proportion of dense neighborhood words then  
33  
34 children are receiving linguistic input that is likely to aid their word learning.  
35  
36  
37

38  
39 A limitation of prior work examining the influence of neighborhood density on child-  
40  
41 directed speech is that previous work has computed neighborhood density using either adult-  
42  
43 directed written language or child-directed speech. When computing neighborhood density  
44  
45 using adult-directed written language (e.g., from a 20,000 word adult dictionary, Storkel,  
46  
47 2009), it is a significant predictor of child vocabularies (Stokes, 2010; Storkel, 2009). When  
48  
49 calculated using child-directed speech (using child-directed speech from Brent & Siskind,  
50  
51 2001), it no longer appears to be (Swingley & Humphrey, 2018), although a network analysis  
52  
53 of children's early productive vocabularies showed that the number of neighbors (calculated  
54  
55 from caregiver utterances) was positively related to when a word enters the child's productive  
56  
57 lexicon (Carlson, Sonderegger, & Bane, 2014). One potential confound across many of these  
58  
59  
60

## Child- and adult-directed speech

8

1  
2  
3 studies is the use of child-directed *speech* versus adult-directed estimates based on *written*  
4 text, the latter being a clear departure from the language that children actually hear. In the  
5 current study we will compare the effect of neighborhood density on child vocabularies when  
6 computing densities based on both child-directed and adult-directed speech.  
7  
8  
9  
10  
11  
12

13 Phonotactic probability has also been shown to hold an influence over children's word  
14 learning. Words that are high in phonotactic probability are those that have common sound  
15 sequences (traditionally measured at the biphone level) within them; whereas words that are  
16 low in phonotactic probability have rarer sound sequences. Storkel (2009) used MCDIs to  
17 show that the receptive vocabularies of young children chiefly consisted of words that were  
18 low in phonotactic probability. Studies using quick incidental learning also show that young  
19 children are more likely to identify a novel object when it is labelled with a novel low  
20 phonotactic probability word than when paired with a novel high phonotactic probability  
21 word (e.g., Storkel & Lee, 2011). That said, once again there is an issue in measure  
22 calculation: the cited work used adult-directed written language corpora for calculations.  
23 When child-directed speech was used, phonotactic probability was not a significant predictor  
24 of early word learning when word frequency was accounted for (Swingley & Humphrey,  
25 2018). We will therefore apply the same corpora for neighborhood density to also examine  
26 phonotactic probability.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

45 In summary, children's early vocabularies contain large numbers of short words, high  
46 frequency words, and low phonotactic probability words, **and a greater proportion of high**  
47 **neighborhood density words, but it is not clear whether this is due to being exposed to more**  
48 **similar distributions in child-directed speech relative to adult-directed speech.**  
49  
50  
51  
52  
53  
54

55 Note that an alternative view could be that the linguistic input has a secondary  
56 influence on children's vocabulary learning. Under this view, the child's own cognitive  
57 apparatus (e.g., phonological short-term memory, Gathercole & Baddeley, 1989) is the  
58  
59  
60

## Child- and adult-directed speech

9

primary constraint on the type of words that can be learned. This suggests that early word learning would be relatively uniform, such as children learning short words regardless of the length of words addressed to them. However, as we will see, children's early word learning shows similar distributions to the linguistic input they hear (see for example Jones et al., 2021), suggesting that the linguistic input plays a large role in vocabulary acquisition.

Our primary focus, therefore, is to examine whether children's learning of words is influenced by the characteristics of the words that they hear. That is, does child-directed speech predominantly contain short words, high frequency words, low phonotactic probability words, and (proportionally) dense neighborhood words; and how predictive are these measures of children's productive vocabularies? Our secondary focus is to also examine adult-directed speech. If, as we argue, child vocabularies are in part driven by the lexical characteristics of words in child-directed speech, then we should expect those characteristics to be more predictive of child vocabularies when they are measured within child-directed speech than when they are measured in adult-directed speech. Examining adult-directed speech will therefore enable us to determine whether word characteristics of adult-directed speech are different to that of child-directed speech and have different predictive powers in relation to child vocabulary learning, since this could add weight to any evidence that children's word learning is influenced by the lexical characteristics of the speech they hear. Additionally, because previous work has shown conflicting results based on the corpora used to calculate word frequency, phonotactic probability, and neighborhood density, computing these using both child-directed and adult-directed speech will show how these differ on the basis of the measurement corpora.

The remainder of this paper is as follows. First, we outline our methodology and the child-directed and adult-directed speech corpora from which our four measures (word length, word frequency, neighborhood density and phonotactic probability) will be derived together

## Child- and adult-directed speech

10

1  
2  
3 with the corpora from which child vocabularies will be derived. We then detail comparisons  
4  
5 across the child-directed and adult-directed speech corpora for all four measures together  
6  
7 with the utility of each corpora in predicting child productive vocabularies (productive rather  
8  
9 than receptive vocabulary is examined because we use child speech corpora rather than  
10  
11 MCDIs). Finally, we discuss our results in the context of current literature.  
12  
13  
14  
15  
16  
17

## Method

## The speech corpora

18  
19  
20  
21  
22 Child-directed speech is aggregated from 6 CHILDES British English corpora,  
23  
24 extracting only the child-directed utterances when the children were 2- to 3-years of age:  
25  
26 Belfast (Henry, 1995), Thomas (Lieven et al., 2009), Tommerdahl (Tommerdahl &  
27  
28 Kilpatrick, 2013), Wells (Wells, 1981), Forrester (Forrester, 2002), and Lara (Rowland &  
29  
30 Fletcher, 2006).  
31  
32

33  
34 Adult-directed speech is taken from the British National Corpus (BNC, 2007). The  
35  
36 BNC is a monolingual British English corpus containing approximately 100 million  
37  
38 orthographic words of spoken (10%) and written language (90%). The spoken BNC was used  
39  
40 in the present study. This consists of transcriptions of unscripted informal conversations  
41  
42 balanced for age, region, context and social class and also formal conversations e.g., from  
43  
44 meetings.  
45  
46

47  
48 Child productive vocabularies were taken from the child utterances of the Manchester  
49  
50 corpus (Theakston, Lieven, Pine, & Rowland, 2001) which involve 12 sets of mother-child  
51  
52 interactions spanning one year between (average) child ages 22-34 months. Recordings for  
53  
54 each mother-child pairing were taken twice every three weeks and orthographically  
55  
56 transcribed. We chose not to use the maternal utterances from this corpora as our child-  
57  
58  
59  
60

1  
2  
3 directed speech to eliminate any overlap in vocabulary use that may influence our analyses  
4  
5 (i.e., the context of mother-child conversations may involve similar vocabulary).  
6  
7

8 For all corpora, compound nouns were separated into their constituent words (e.g.,  
9  
10 *merry-go-round* as three words: *merry*, *go*, and *round*). The rationale for this was: (1) our  
11  
12 corpus analysis involves 2- to 3-year-old children who are able to both produce novel  
13  
14 compounds and are able to determine that words in compounds have different roles (e.g.,  
15  
16 interpreting cheese+knife as a knife to cut cheese rather than a knife made out of cheese),  
17  
18 suggesting children of this age understand the individual words in compounds (Clark,  
19  
20 Gelman & Lane, 1985); (2) our corpus analyses rely on transcriptions of compounds that vary  
21  
22 within and across transcribers (e.g., Christmas tree and Christmas+tree are transcribed), so  
23  
24 separating out compounds make our analyses more consistent; and (3) many compounds,  
25  
26 such as *merry-go-round*, would have zero neighbors if treated as one entity. To show that the  
27  
28 way in which we treat compounds does not influence our results, in Appendix 1 we duplicate  
29  
30 our analyses when excluding all compounds, with the results being identical to those in the  
31  
32 main paper.  
33  
34  
35  
36  
37

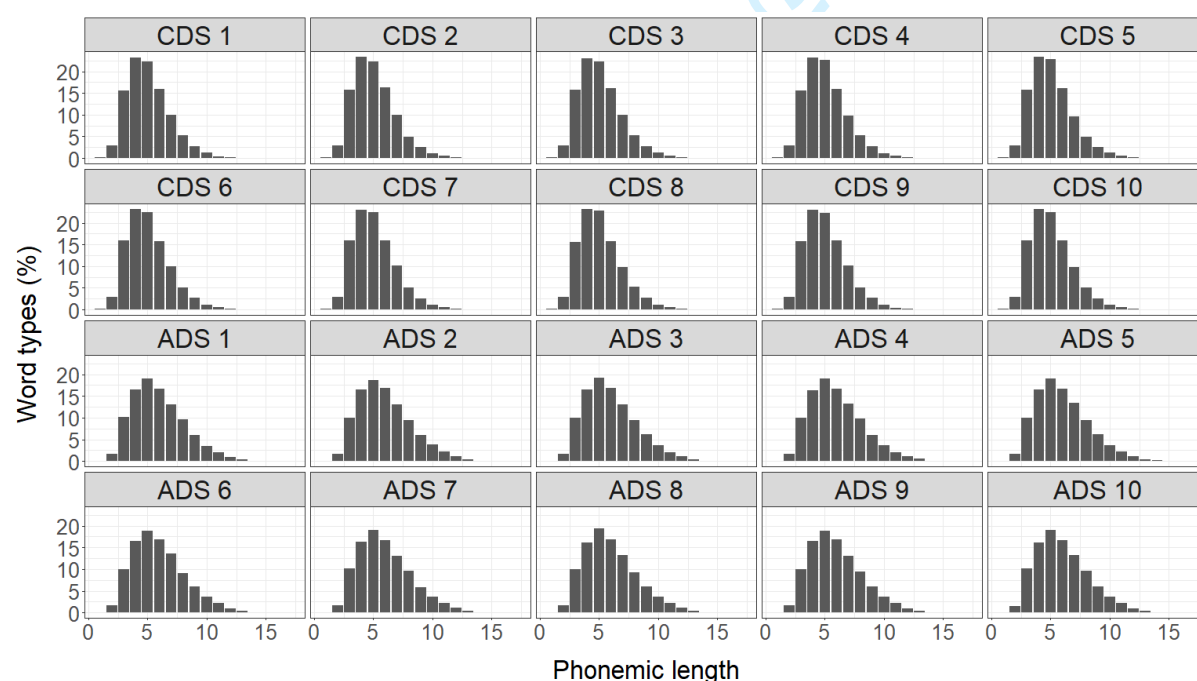
38 Given that the corpora involve orthographic transcriptions and three of our four  
39  
40 measures require phonological transcriptions, we converted all orthographic forms to  
41  
42 phonological forms using the CMU Pronouncing Dictionary  
43  
44 (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>) containing over 134,000 American English  
45  
46 phonetic transcriptions of words. Commonly occurring words in the children's speech  
47  
48 corpora that were not present in the dictionary were manually added together with their  
49  
50 phonemic equivalent, as were words that were transcribed differently to how they were  
51  
52 spoken (e.g., won't in place of willn't). Orthographic utterances were ignored when they  
53  
54 contained a word that did not have a corresponding entry in the dictionary. Unless otherwise  
55  
56 stated, all quantities reported for corpora are based on the phonological form. Note that by  
57  
58  
59  
60

## Child- and adult-directed speech

12

1  
2  
3 considering the phonological form only, some orthographically different word types (e.g., *to*,  
4 *two*, *too*) are treated as one word type.  
5  
6

7  
8 The complete set of children's phonemically-transcribed utterances comprised  
9  
10 531,940 word tokens, whereas the child-directed speech comprised 1,431,271 word tokens  
11  
12 and the adult-directed speech comprised 6,974,576 word tokens. Differences in sample size  
13  
14 are problematic for comparisons across corpora because distributional properties change as  
15  
16 sample size increases - in particular, increases in number of word types decline as sample size  
17  
18 increases (see Montag, Jones, & Smith, 2018). We therefore created ten random samples of  
19  
20 the child-directed speech and adult-directed speech to each match the children's utterances  
21  
22 for number of word tokens. This was accomplished by randomly selecting utterances from  
23  
24 the child-directed speech and adult-directed speech corpora until the number of word tokens  
25  
26 matched that of the children. The distributions of word length, word frequency, phonotactic  
27  
28 probability and neighborhood density were very similar across all child-directed speech  
29  
30 samples and across all adult-directed speech samples (see Figure 1 for word length  
31  
32 distributions across different random samples) and we therefore used one random sample  
33  
34 from each for subsequent reporting and analyses.  
35  
36  
37  
38  
39



## Child- and adult-directed speech

13

1  
2  
3 Figure 1. Distribution of word lengths across ten random samples from the child-directed  
4 speech (upper two panels, labelled CDS 1-10) and adult-directed speech (lower two panels,  
5  
6 labelled ADS 1-10) to illustrate that the distributions from each sample are almost identical.  
7  
8  
9

10  
11  
12 Vocabulary examination

13  
14 Data manipulation and analysis were carried out using the R programming language (version  
15  
16 3.5.2; R Core Team, 2018).  
17  
18

19 All corpora included tags to indicate noun plural and so for all word type analyses,  
20  
21 noun plurals were reduced to their singular form (e.g., *car* for *cars*).  
22  
23

24 Calculations were produced for each of our four measures. Phonemic word length  
25  
26 does not change based on the corpora used and was the phonemic length of each word type.  
27  
28 The remaining three measures were computed for both child-directed speech and adult-  
29  
30 directed speech.  
31  
32

33 Word frequency was the number of occurrences of each word type within the  
34  
35 respective corpora, computed as log<sub>10</sub> frequencies.  
36  
37

38 Neighborhood density was defined as stress-unmarked phonological neighborhood  
39  
40 density, referring to the unweighted count of words that differed from a given word by one  
41  
42 phoneme (i.e., neighborhood density was a count of all word types that were neighbors of the  
43  
44 target word in the corpora, in line with the majority of studies involving neighborhood  
45  
46 density).  
47  
48

49 Phonotactic probability was defined as stress-unmarked word-average biphone  
50  
51 probability, referring to the weighted likelihood of the occurrence of ordered phoneme pairs  
52  
53 that were present in a given word (i.e., phonotactic probability reflected how often the  
54  
55 biphone occurs in the corpora, in line with the majority of studies involving phonotactic  
56  
57 probability). This was computed by the second author using the formulae from the Irvine  
58  
59  
60

## Child- and adult-directed speech

14

Phonotactic Online Dictionary (version 2.0; IPhOD, Vaden, Halpin, & Hickok, 2009, see <http://www.iphod.com/>) obtained from the main author of the dictionary (Vaden, personal correspondence, 21/07/2019).

All words, together with their lexical measures as computed from the child-directed corpora and as computed from the spoken BNC can be downloaded here:

<https://www.ntu.ac.uk/research/groups-and-centres/groups/language-literacy-and-psycholinguistics>.

## Results

All analyses scripts can be found here: <https://doi.org/10.17605/OSF.IO/TCQU5>. We first outline the descriptive statistics for each corpus and then provide analyses for phonemic word length, word frequency, neighborhood density, and phonotactic probability. For our analyses, three of the four aforementioned measures (word frequency, phonotactic probability, and neighborhood density) are computed using child-directed speech and adult-directed speech, since these may vary depending on the particular corpora used to compute them. Note that in these analyses, all word categories are considered. Please see the Appendix 2 for plots that are broken down by word category (noun, verb, adjective, and ‘other’), which show that there is little difference when considering all word categories as a whole versus when breaking them down by category.

## Descriptive statistics for corpora

Table 1 shows descriptive statistics for child productive vocabularies, child-directed speech, and adult-directed speech. There are two notable aspects: (1) by examining children’s productive vocabularies (4,051 phonological word types), we capture many more words than one would if using CDIs (~ 650 words); and (2) the number of word types increases with the age at which speech is directed (i.e., child-directed vs. adult-directed).



Table 1. Descriptive statistics for phonological word types/tokens for each corpora (child productive vocabularies, child-directed speech, adult-directed speech).

Corpora	Total word types	Total word tokens	Sample word types	Sample word tokens
Child vocabularies	4,051	531,940	N/A	N/A
Child-directed speech	8,573	1,431,271	6,503	531,940
Adult-directed speech	29,539	6,974,576	13,942	531,940

Word length

As one may expect, child-directed speech is characterized by the use of shorter word types on average than adult-directed speech (see Figure 2). Table 2 also shows that child productive vocabularies contain proportionally more short word types than child-directed speech and in turn, child-directed speech contains proportionally more short word types than adult-directed speech.

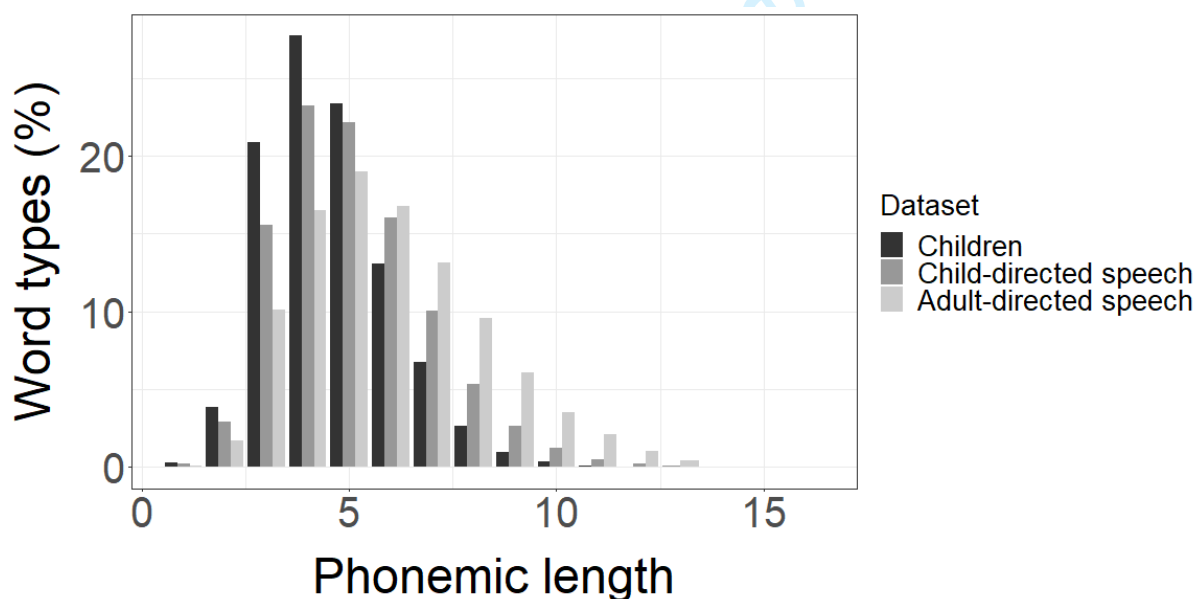


Figure 2. Percentage of word types varying in phonemic length, for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech).

Table 2. Kolmogorov-Smirnov comparisons for phonemic word length between different corpora, with 95% bootstrapped confidence intervals (2.5% and 97.5% percentiles, 1,000 iterations). *p* values and confidence intervals were corrected using Holm's correction.

Comparison	D	<i>p</i>	2.5%	97.5%
Children vs. Child-directed speech	.12	< .001	.10	.14
Child- vs. Adult-directed speech	.17	< .001	.15	.18
Children vs. Adult-directed speech	.29	< .001	.27	.31

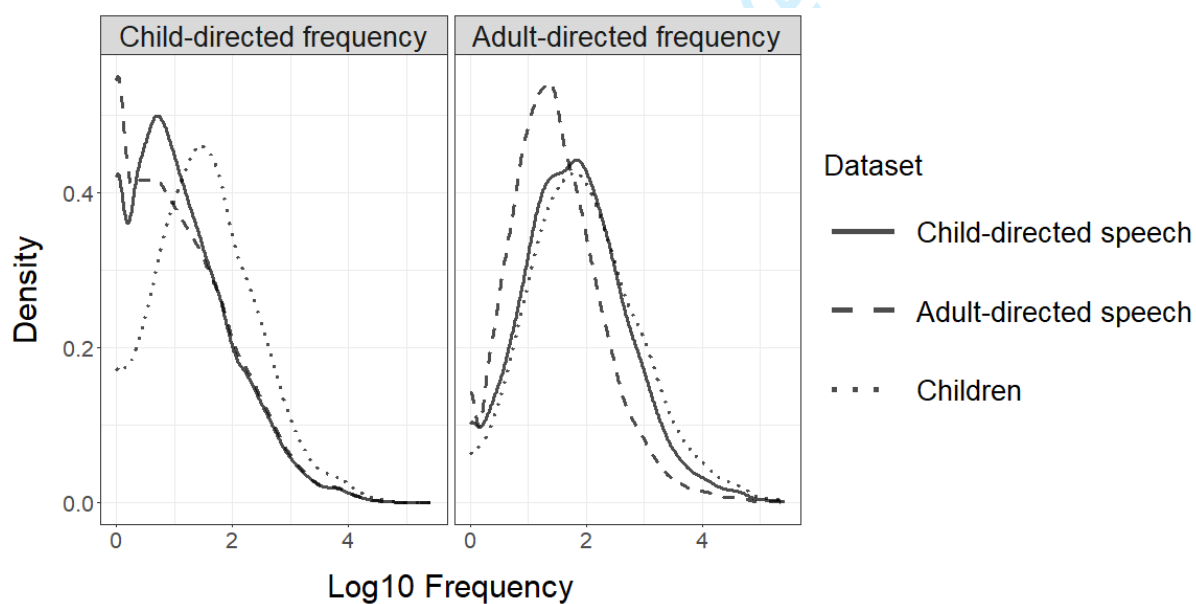
### Word frequency

Figure 3 shows how the proportion (plotted as density) of word types vary by word frequency across the three corpora when computing word frequency based on child-directed speech and adult-directed speech. Regardless of the corpora used to compute frequencies, children produce greater numbers of high frequency word types relative to those that appear in child-directed speech, and in turn, child-directed speech contains greater numbers of high frequency word types than adult-directed speech. These effects are confirmed by Kolmogorov-Smirnov comparisons (see Table 3). Of note is that the density plots in Figure 3 show that the frequencies of words used by children and in child-directed speech are much closer to one another than those involving adult-directed speech – regardless of which corpora are used to compute word frequency – suggesting that children's productive vocabularies are influenced by the frequency of the words they hear in child-directed speech<sup>3</sup>.

<sup>3</sup> Note that the number of word types within each corpus directly influences word frequency: adult-directed speech has many more word types than the children and child-directed speech (see Table 1) which inevitably means that adult-directed speech will have greater numbers of low frequency words.

Table 3. Kolmogorov-Smirnov comparisons for Log10 word frequency between different corpora, with 95% bootstrapped confidence intervals (2.5% and 97.5% percentiles, 1,000 iterations). Comparisons are performed when computing Log10 word frequency using child-directed speech and adult-directed speech.

Comparison	Word frequency				
	measure	D	<i>p</i>	2.5%	97.5%
Children vs. Child-directed speech	Child-directed	.23	< .001	.21	.26
Child- vs. Adult-directed speech	Child-directed	.05	< .001	.04	.07
Children vs. Adult-directed speech	Child-directed	.24	< .001	.22	.27
Children vs. Child-directed speech	Adult-directed	.06	< .001	.05	.09
Child- vs. Adult-directed speech	Adult-directed	.18	< .001	.17	.20
Children vs. Adult-directed speech	Adult-directed	.23	< .001	.22	.26



## Child- and adult-directed speech

18

1  
2  
3 Figure 3. Density plot showing how often words occur across the frequency range in each  
4 sample (child productive vocabularies, child-directed speech, adult-directed speech), with  
5  $\text{Log}_{10}(\text{frequency})$  computed using the child corpus (left panel) and adult corpus (right panel).  
6  
7  
8  
9  
10 For example, the peak of the dotted line in the right panel is at a  $\text{Log}_{10}(\text{frequency})$  of just  
11 below 2, indicating that the words children produce most often have a  $\text{Log}_{10}(\text{frequency})$  of  
12 just below 2 when using the adult corpus to measure frequency. Note that because the  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Neighborhood density

Figure 4 shows how the proportion (plotted as density) of words produced by children and used in child-directed and adult-directed speech vary in neighborhood density, when calculating neighborhoods using child-directed speech and adult-directed speech. In general, child productive vocabularies contain word types from more dense neighborhoods than child-directed speech (i.e., lower proportions of low neighborhood density words for child productive vocabularies relative to child-directed speech but higher proportions of higher density neighborhood words), and child-directed speech contains word types from more dense neighborhoods than adult-directed speech (see also, statistical comparisons, Table 4). However, Figure 4 clearly shows at least two things of note: (a) differences across children, child-directed speech and adult-directed speech mainly stem from **differences in the quantities** of low neighborhood words in children's productive vocabularies (and slightly lower amounts of low neighborhood words in child-directed speech); and (b) differences between child-directed speech and adult-directed speech are more pronounced when neighborhood is calculated using the adult-directed corpora. At this stage it is not clear whether the neighborhood densities of words used in child-directed speech influence child vocabularies

relative to adult-directed speech but we will return to this when examining predictors of children’s productive vocabulary.

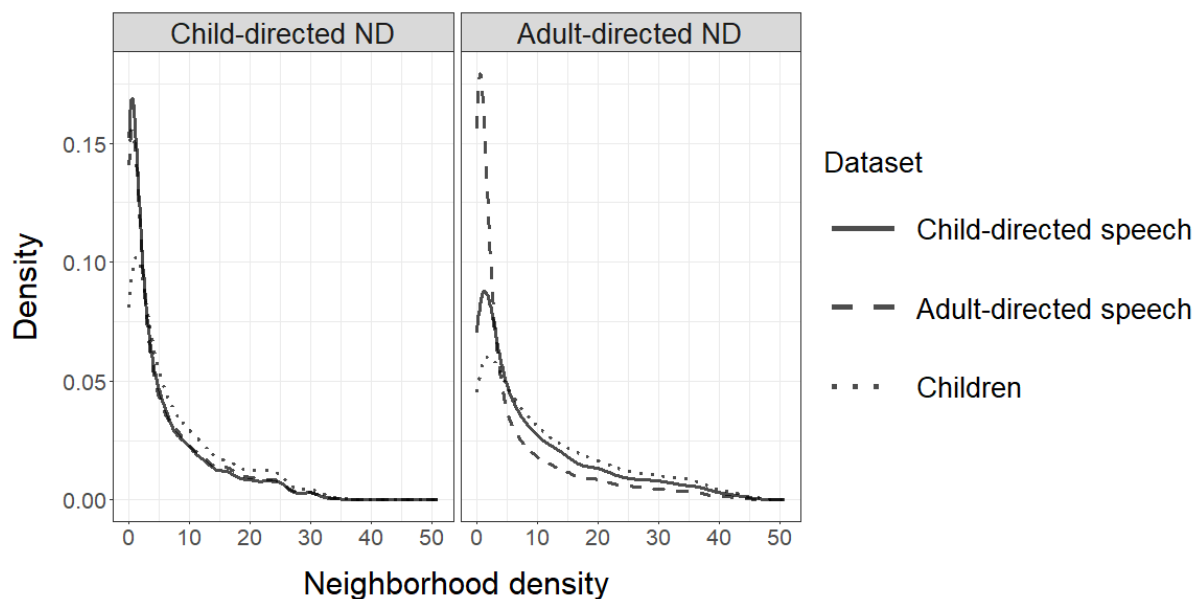


Figure 4. Density plot showing how often words occur across the neighborhood density range in each sample (child productive vocabularies, child-directed speech, adult-directed speech), with neighborhood density computed using the child corpus (left panel) and adult corpus (right panel). Regardless of sample used, the peaks are always just above 0, indicating that a low neighborhood density is most frequent across all samples.

Table 4. Kolmogorov-Smirnov comparisons for neighborhood density between different corpora, with 95% bootstrapped confidence intervals (2.5% and 97.5% percentiles, 1,000 iterations). Comparisons are performed when computing neighborhood density using child-directed speech and adult-directed speech.

Comparison	Neighborhood		D	p	2.5%	97.5%
	density measure					

## Child- and adult-directed speech

Children vs. Child-directed speech	Child-directed	.13	< .001	.11	.16
Child- vs. Adult-directed speech	Child-directed	.02	.083	.01	.04
Children vs. Adult-directed speech	Child-directed	.13	< .001	.10	.15
Children vs. Child-directed speech	Adult-directed	.11	< .001	.09	.14
Child- vs. Adult-directed speech	Adult-directed	.19	< .001	.17	.21
Children vs. Adult-directed speech	Adult-directed	.30	< .001	.27	.32

## Phonotactic probability

Figure 5 shows how the proportion (plotted as density) of word types vary by phonotactic probability for children, child-directed speech and adult-directed speech, for each corpora used to calculate phonotactic probability. The figure shows that children produce greater numbers of low phonotactic probability words relative to those that appear in child-directed speech, and in turn, child-directed speech contains greater numbers of low phonotactic probability word types than adult-directed speech. These conclusions are generally confirmed by Kolmogorov-Smirnov and confidence interval analyses (see Table 5). That said, one can see from Figure 5 that any differences are smaller when phonotactic probability is computed using child-directed speech than it is when computed using adult-directed speech. As with neighborhood density, we will examine this further when analyzing predictors of children's productive vocabularies.

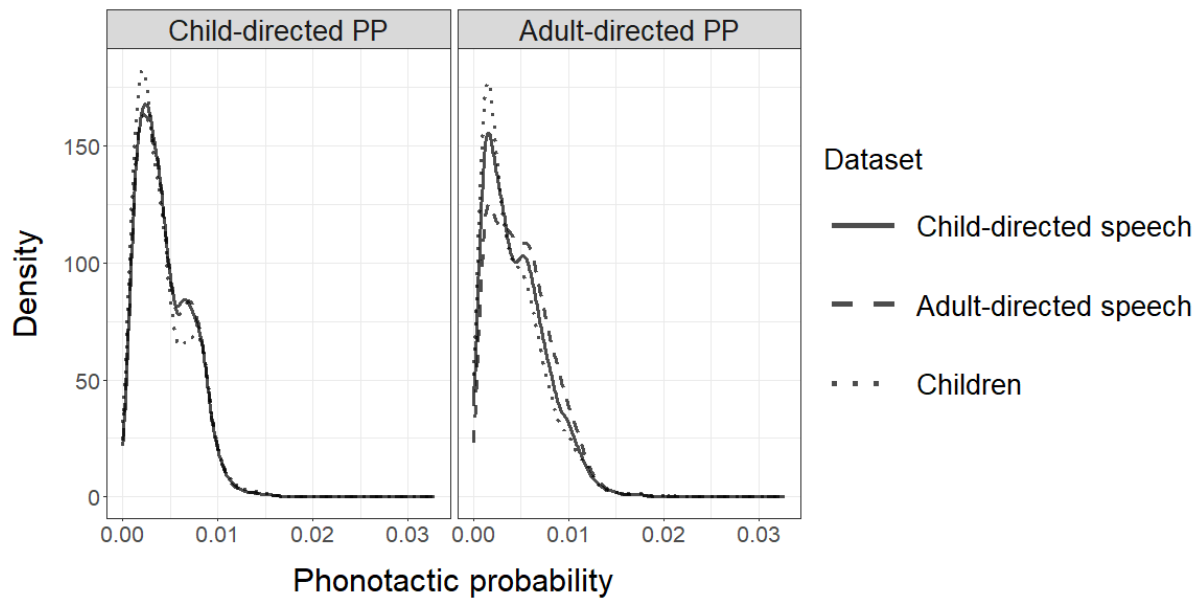


Figure 5. Density plot showing how often words occur across the phonotactic probability range in each sample (child productive vocabularies, child-directed speech, adult-directed speech), with phonotactic probability computed using the child corpus (left panel) and adult corpus (right panel). Regardless of sample used, the peaks are always just above 0, indicating that a low phonotactic probability is most frequent across all samples.

Table 5. Kolmogorov-Smirnov comparisons for phonotactic probability between different corpora, with 95% bootstrapped confidence intervals (2.5% and 97.5% percentiles, 1,000 iterations). Comparisons are performed when computing phonotactic probability using child-directed speech and adult-directed speech.

Comparison	Phonotactic probability measure	Phonotactic			
		D	<i>p</i>	2.5%	97.5%
Children vs. Child-directed speech	Child-directed	.05	< .001	.04	.08
Child- vs. Adult-directed speech	Child-directed	.01	.977	.01	.03
Children vs. Adult-directed speech	Child-directed	.05	< .001	.03	.08
Children vs. Child-directed speech	Adult-directed	.06	< .001	.04	.08

## Child- and adult-directed speech

22

Child- vs. Adult-directed speech	Adult-directed	.09	< .001	.07	.11
Children vs. Adult-directed speech	Adult-directed	.14	< .001	.12	.16

Up to this point, we have examined our four measures independently, showing how each measure differs between child-directed speech and adult-directed speech. However, these measures interact with one another. For example, as words get longer, both their frequency of occurrence and neighborhood density decline (see Jones et al., 2021 for discussion). In addition, while we have examined whether there are differences in the measures across child-directed and adult-directed speech, we do not know if these differences affect children's vocabulary learning. In the next section we therefore examine whether the measures – together with pairwise interactions – are predictive of children's productive vocabularies.



Word length, word frequency, neighborhood density and phonotactic probability as predictors of child productive vocabularies

We now examine whether our four measures and their pairwise interactions are predictive of child productive vocabularies on the basis of which sample the measures are computed from i.e., whether child-directed speech is more predictive of child productive vocabularies than adult-directed speech. In order to do this, we need to see whether there is anything particular in relation to our four measures regarding words that the children produce versus words that they fail to produce. We therefore capitalized on the Manchester corpus containing both maternal and child utterances that are part of ongoing caregiver-child conversations. To make our analyses tractable, we only considered nouns. All nouns appearing in child utterances were classified as *produced* nouns. All nouns appearing in maternal utterances but not in the child utterances were classified as *not produced*, since we have no evidence of any of the children speaking those nouns. A major advantage of using the maternal utterances that are part of the ongoing mother-child conversations is that all nouns have been heard by the children, regardless of whether they subsequently produced the nouns. This method also allows us to examine whether there are any particular features (across our four measures and their pairwise interactions) of produced nouns that distinguish them from those that are not produced. Table 6 shows the sample size of produced and not produced nouns, by phonemic length<sup>4</sup>. Produced nouns outweigh not produced nouns for words up to five phonemes in length, but beyond that there are more nouns that children fail to produce than those that they produce. In subsequent analyses, nouns of length greater than 8 phonemes were discarded due to low sample sizes.

---

<sup>4</sup> Note that our analyses examine lexical characteristics but also the variance explained by child, and so the produced nouns in Table 6 are totals across the 12 children in the corpus (e.g., for ‘truck’, the produced nouns column would include a score for every child who produced ‘truck’). This explains

## Child- and adult-directed speech

Table 6. Number of nouns that are produced by children together with those that appear in maternal utterances but not in the child utterances (i.e., not produced by children), by phonemic length.

Phonemic length	Produced	Not produced	Ratio
2	335	95	3.53
3	2,423	1,092	2.22
4	2,261	1,426	1.59
5	1,666	1,287	1.29
6	725	808	0.90
7	449	588	0.76
8	195	287	0.68
9	52	128	0.41
10	13	79	0.16
11	5	29	0.17

---

why the produced noun totals are greater than the 4,051 words in Table 1 which looks at the corpora as a whole and are not restricted to nouns.

1  
2  
3 Figure 6 shows the distribution of each of the four measures used in the current study,  
4 for produced and not produced nouns when the measures are computed using child-directed  
5 speech and adult-directed speech. The figure largely confirms our previous findings whereby  
6 children's productive vocabularies are dominated by short words, high frequency words, low  
7 phonotactic probability words and **proportionally greater numbers of words from dense**  
8 **neighborhoods** – regardless of the source used to compute the measures. That said, the  
9 evidence for phonotactic probability is marginal. The data also show that all of the measures  
10 are skewed. For analysis, we transformed (standardized) the data so that odds ratios can be  
11 directly compared. Inspection of the model's residuals using the DHARMA package (Hartig,  
12 2022) indicated no problems relating to deviation from the expected distribution of the data.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

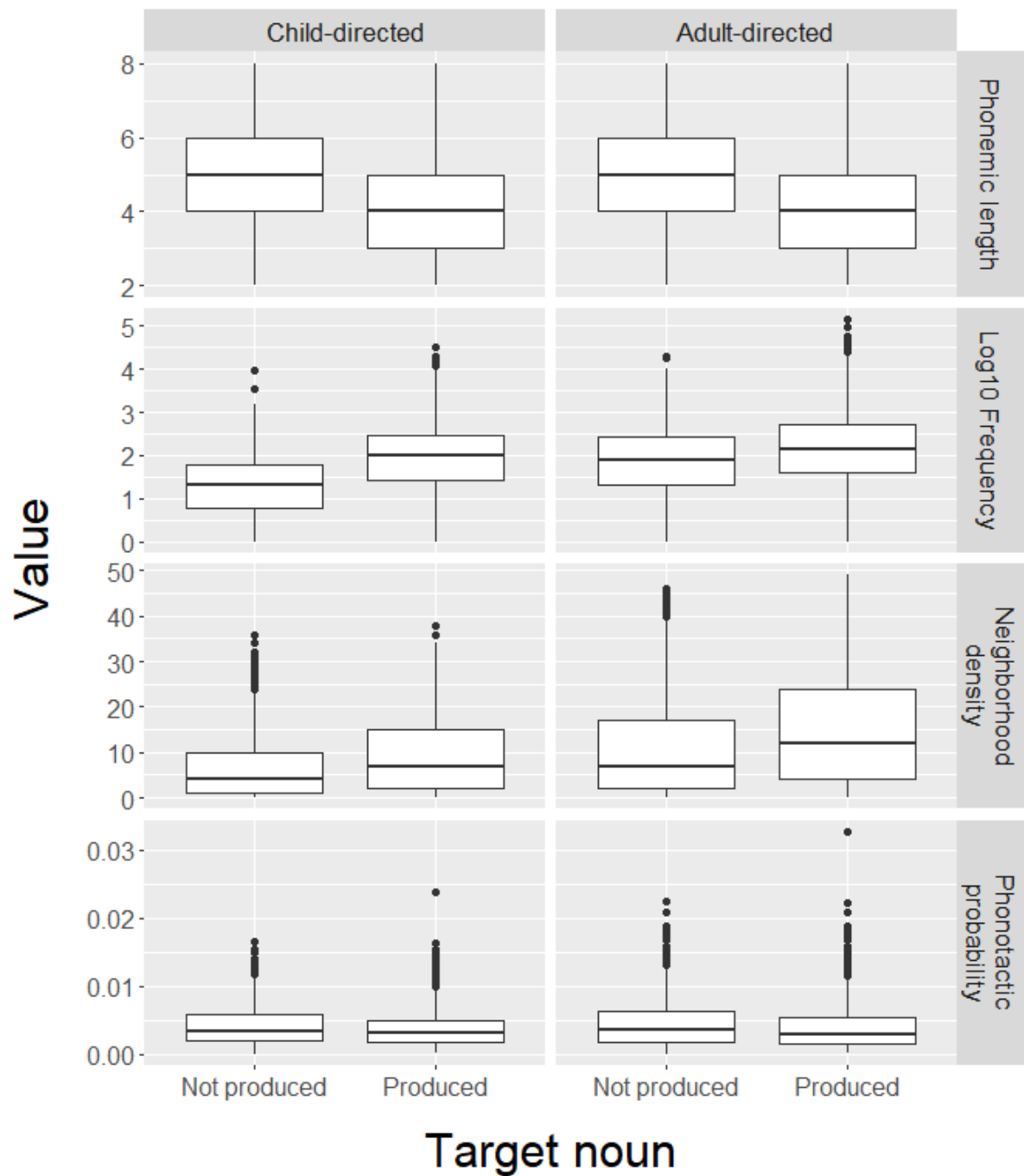


Figure 6. Box plot comparisons of nouns produced by children in the Manchester corpus (produced) and nouns appearing in the maternal utterances of the corpus that were not produced by children (not produced) for each of the four measures that are the focus of the current study, with each measure computed using the child-directed speech and adult-directed speech corpora.

1  
2  
3 Table 7 provides mixed effects logistic regression analyses using our four measures as  
4  
5 predictors of whether a noun was produced or not, when computing the measures using the  
6  
7 child-directed speech and adult-directed speech corpora. There are at least three notable  
8  
9 aspects of this analysis that favor child-directed speech over adult-directed speech. First,  
10  
11 child-directed speech explains over twice the variability in child productive vocabulary than  
12  
13 adult-directed speech (Conditional  $R^2$  of 0.263 vs. 0.124). Second, although both child-  
14  
15 directed and adult-directed speech show word length and word frequency as significant  
16  
17 predictors (with neighborhood density and phonotactic probability notably absent), the odds  
18  
19 ratios<sup>5</sup> indicate the effect of word length is similar across the two different corpora whereas  
20  
21 word frequency has a much greater influence when measured by child-directed speech than  
22  
23 adult-directed speech. Third, there is an additional interaction effect for child-directed speech  
24  
25 compared to adult-directed speech. For completeness, Figure 7 investigates all pairwise  
26  
27 interaction effects, though the key interactions from Table 7 are those between neighborhood  
28  
29 density and frequency and neighborhood density and phonotactic probability. The most  
30  
31 influential based on odds ratios is between frequency and neighborhood density, showing that  
32  
33 while frequency influences child noun production, it is particularly so when the nouns in  
34  
35 question are also high in neighborhood density. The neighborhood density and phonotactic  
36  
37 probability interaction shows that as phonotactic probability increases, neighborhood density  
38  
39 has an increasingly influential effect: children are much more likely to produce nouns having  
40  
41 a high phonotactic probability when those nouns are high in neighborhood density. In short,  
42  
43 the interactions show that neighborhood density is an important factor, particularly when  
44  
45 measured using child-directed speech.  
46  
47  
48  
49  
50  
51  
52  
53

---

58  
59 <sup>5</sup> For odds ratios, the greater the deviation from 1.00, the more the predictor wields an influence on  
60 the outcome measure.

## Child- and adult-directed speech

Table 7. Mixed effects logistic regression analyses predicting word production of nouns (levels: produced/not produced) as a function of the four continuous predictors (phonemic length, Log10 word frequency, neighborhood density, phonotactic probability), and the pairwise interactions of those predictors. Random intercept of child is included. Child predictors are computed using standardized child-directed speech and adult-directed speech measures, so that odds ratios can be compared. Note: neighborhood = neighborhood density; phonotactic = phonotactic probability.

<i>Predictors</i>	<b>Child-directed model</b>			<b>Adult-directed model</b>		
	<i>Odds Ratios</i>	<i>CI</i>	<i>P</i>	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	1.78	1.43 – 2.22	<0.001	1.52	1.22 – 1.89	<0.001
length	0.83	0.75 – 0.93	0.001	0.78	0.71 – 0.86	<0.001
frequency	2.47	2.35 – 2.59	<0.001	1.34	1.28 – 1.39	<0.001
neighborhood	1.03	0.91 – 1.16	0.654	1.08	0.98 – 1.19	0.134
phonotactic	0.99	0.94 – 1.03	0.556	1.01	0.97 – 1.05	0.683
length * neighborhood	0.95	0.86 – 1.04	0.242	1.00	0.92 – 1.07	0.904
length * phonotactic	0.98	0.92 – 1.05	0.586	0.97	0.91 – 1.03	0.290
length * frequency	1.01	0.94 – 1.09	0.727	1.00	0.94 – 1.07	0.988
frequency * neighborhood	1.16	1.08 – 1.25	<0.001	1.23	1.15 – 1.31	<0.001
frequency * phonotactic	0.97	0.92 – 1.02	0.237	0.96	0.92 – 1.00	0.070
neighborhood * phonotactic	1.09	1.02 – 1.17	0.015	1.06	0.99 – 1.13	0.076
<b>Random Effects</b>						
$\sigma^2$	3.29			3.29		
$\tau_{00}$	0.13 <sub>id</sub>			0.14 <sub>id</sub>		
ICC	0.04			0.04		
N	12 <sub>id</sub>			12 <sub>id</sub>		
Observations	12299			13328		
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.234 / 0.263			0.087 / 0.124		

Child- and adult-directed speech

29

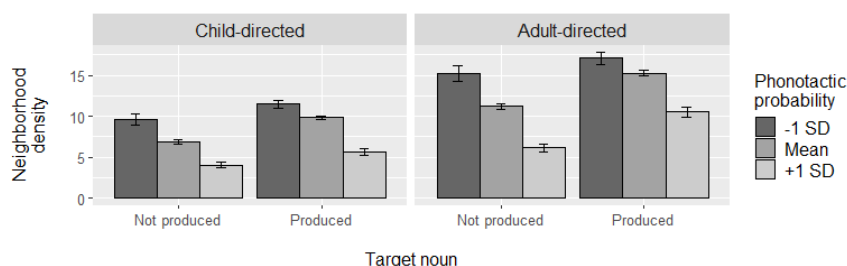
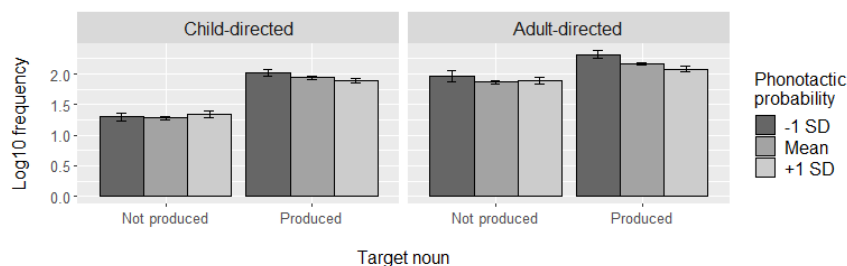
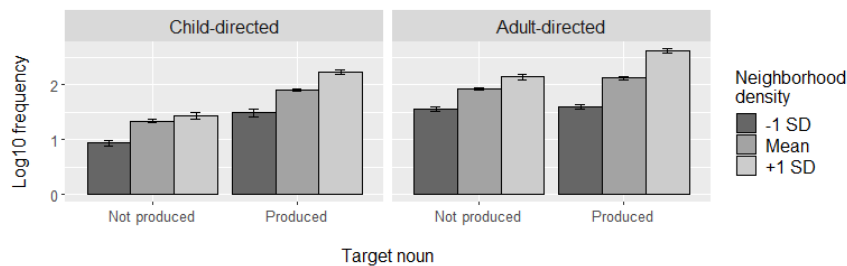
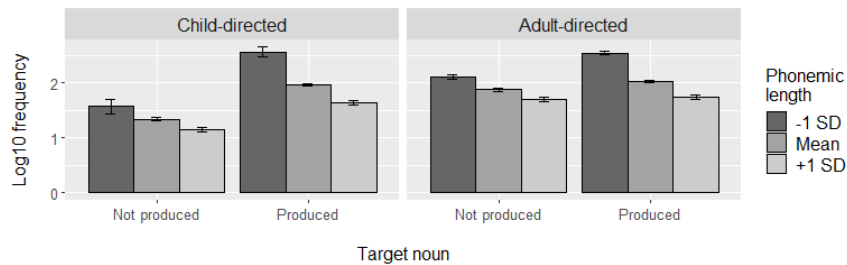
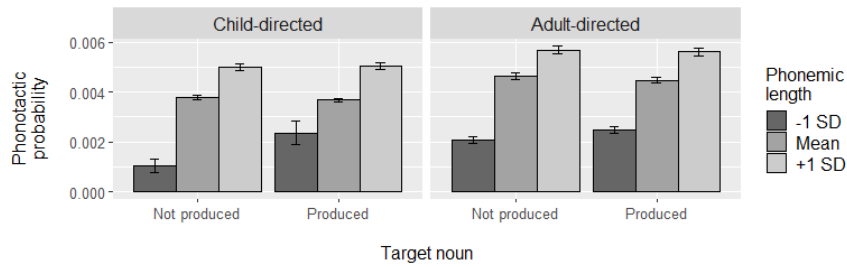
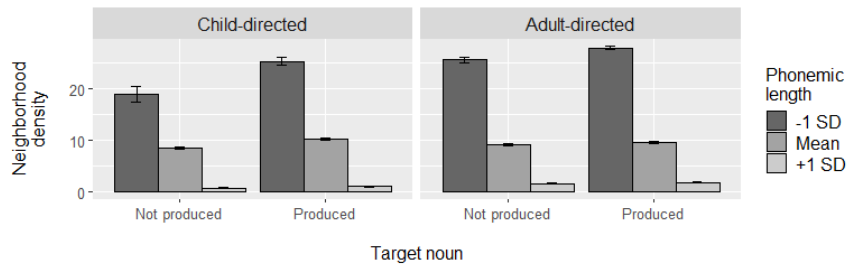


Figure 7. Box plot comparisons of pairwise interactions across word length, log10 frequency, neighborhood density and phonotactic probability, when these measures are computed using either child-directed speech or adult directed speech. Error bars indicate confidence intervals.

## Discussion

This paper set out to examine whether the lexical characteristics of children's early vocabularies might be driven by the lexical characteristics of the input they receive. We compared large-scale samples of child-directed speech to the same quantity of adult-directed speech, finding that a plausible reason that children's vocabularies contain short words, high frequency words, words **that are low in phonotactic probability, and words that are proportionally high in neighborhood density**, is that child-directed speech contains greater proportions of words that fit these categories than adult-directed speech. Regression analyses then examined the extent to which these factors – measured using either child-directed speech or adult-directed speech – were predictive of children's productive vocabularies. Child-directed speech explained over twice the variability in children's noun productions than did adult-directed speech, with the frequency of encounter of a noun in child-directed speech being a key factor and word length a secondary factor. However, there were also two interaction effects relating to how child-directed speech influenced children's productive vocabularies, both involving neighborhood density. The lexical characteristics of the words used in child-directed speech clearly play an important role in children's vocabulary acquisition. We will first consider the significant predictors relating to child-directed speech and their implications for children's vocabulary acquisition and then consider what the differences in lexical characteristics between child-directed speech and adult-directed speech may mean.



1  
2  
3 It is by no means surprising that frequency plays a vital role in what words children  
4 learn from child-directed speech. Swingley and Humphrey (2018) found frequency of  
5 occurrence in child-directed speech to be far and away the most consistent predictor of  
6 receptive and productive vocabularies in 15-month old infants; child-directed speech is  
7 littered with repetitions that push frequency of occurrence upwards (Hills, 2013); and  
8 quantity of speech in early childhood is a better predictor of vocabulary than the diversity of  
9 vocabulary (Rowe, 2012; Jones & Rowland, 2017). Our child-directed and adult-directed  
10 speech samples vastly differed in the number of phonological word types (8,573 vs. 29,539).  
11 Given the samples were matched for word tokens, this means greater frequency, and hence  
12 greater repetition, of many word types in child-directed speech relative to adult-directed  
13 speech, further supporting the aforementioned literature. Children also seem to find it easier  
14 to learn short words over long words, which is also not altogether surprising since almost  
15 every learning mechanism follows this pattern (e.g., PARSER: Perruchet & Vinter, 1998;  
16 TRACX: French, Addyman & Mareschal, 2011). Yet frequency and word length cannot be  
17 the only key factors in early word learning. They have to give way to other developmental  
18 factors, not least because at the tail end of the developmental continuum – adult-directed  
19 speech – there is a large amount of infrequent words and long words (see Figures 2 and 3).  
20 This means that other lexical characteristics have to come into play at the expense of  
21 frequency and length as development progresses. The beginnings of this transition are  
22 arguably present in the current data.

23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
There are two effects, both involving neighborhood density, that warrant some  
scrutiny. Neighborhood density influences the effects of both frequency and phonotactic  
probability, with frequent words and words containing biphone pairs that occur frequently  
being more likely to be produced by children if the words also have many similar sounding  
other words in child-directed speech. While there are numerous studies that have shown

## Child- and adult-directed speech

1  
2  
3 neighborhood density to be linked to vocabulary acquisition (e.g., Stokes, 2010; Storkel,  
4  
5 2009), there are others that show little or no effect (e.g., Swingley & Humphrey, 2018). One  
6  
7 plausible explanation for this is that neighborhood effects may emerge later in development,  
8  
9 with those studies showing an effect of neighborhood examining child ages of around 24  
10  
11 months and those showing no effect examining child ages of around 15 months. In addition,  
12  
13 many studies examine receptive vocabulary; given that production lags behind  
14  
15 comprehension (and production may also be more influenced by neighborhood than  
16  
17 comprehension given there is an articulatory factor), our production data (involving an  
18  
19 average child age of 28 months) are consistent with the hypothesis that neighborhood is an  
20  
21 effect that appears at around two years of age.  
22  
23  
24  
25  
26

27  
28 However, one has to explain why neighborhood may influence vocabulary  
29  
30 acquisition. In and of itself, the fact that child-directed speech contains proportionally more  
31  
32 words with many neighbors than adult-directed speech does not tell us why children benefit.  
33  
34 In order for children to realize such benefits, they have to know something about the  
35  
36 neighboring words – so children must learn something about the neighbors themselves, heard  
37  
38 in the child-directed input, and use this information to their advantage. One possibility  
39  
40 (Walley, 1993; Walley, Metsala & Garlock, 2003) is that children initially represent words as  
41  
42 unanalyzed wholes, but by hearing many similar sounding words, come to realize that they  
43  
44 need to learn segmental detail for such words. Such segmental detail presumably facilitates  
45  
46 the learning of other similar sounding, neighboring words. An alternative but not inconsistent  
47  
48 view (Jones et al., 2021) is that children learn bottom-up from exposure to linguistic input,  
49  
50 with greater information being learned about the constituent sequences of sounds within a  
51  
52 word as exposure to that word increases. On this view, children learn sound sequences within  
53  
54 the words they hear. Should some of those words be neighbors of a to-be-learned word, then  
55  
56 the overlap in learned phonological knowledge from neighbors of the new word will scaffold  
57  
58  
59  
60

1  
2  
3 learning of the new word. The traditional definition of neighbor is therefore dispensed with  
4  
5 since neighboring words are those that have some overlap in the sequence of sounds that  
6  
7 comprise the words. The added advantage of Jones et al.'s (2021) view, is that it also explains  
8  
9 why both word length and word frequency are influential in children's word learning – since  
10  
11 gradual learning is based on exposure to words, short words and frequent words are learned  
12  
13 more quickly. In the former case, this is because short words require fewer exposures to be  
14  
15 learned than do long words. In the latter case it is because frequency is directly linked to how  
16  
17 quickly a word is learned, with greater exposure to the word leading to faster learning of it.  
18  
19  
20  
21

22  
23 The hypothesis put forward by Jones et al. (2021) also makes sense of the interaction  
24  
25 effects seen in the current data. Frequency effects are stronger when there are many words  
26  
27 that share sequences of sounds compared to only a few; this is because knowledge of the  
28  
29 constituent phoneme sequences in the other words can be used when learning the new word  
30  
31 such that fewer exposures are required. The interaction between phonotactic probability and  
32  
33 neighborhood, whereby learning high phonotactic probability words is helped by the words  
34  
35 having high neighborhoods, is also readily explained. Having frequent sound sequences that  
36  
37 also appear across many words increases the chances of phonological knowledge being  
38  
39 learned for those sound sequences, and in turn this increases the chances of learning the word  
40  
41 itself.  
42  
43  
44  
45

46  
47 We now turn to the contributions of child-directed speech and adult-directed speech in  
48  
49 children's vocabulary acquisition. One of the key findings of the current study is that there  
50  
51 are differences between child-directed speech and adult-directed speech on every one of the  
52  
53 four key measures used. This has implications for those studies that have examined children's  
54  
55 word learning using measures that are derived from adult-directed speech or adult-directed  
56  
57 written literature. If child-directed speech differs from adult-directed speech on all measures,  
58  
59 one can expect similar differences for adult-directed written literature. Although previous  
60

## Child- and adult-directed speech

1  
2  
3 studies using adult-directed written literature to estimate word frequency, phonotactic  
4 probability, and neighborhood density (e.g., Stokes, 2010; Storkel, 2009) are broadly  
5  
6 consistent with the results presented here when using child-directed speech corpora, they are  
7  
8 likely under-estimating the influence of these measures.  
9  
10  
11  
12

13 Our data show that there are key aspects of child-directed speech – word frequency,  
14 word length, and neighborhood density – that influence child productive vocabularies. In  
15 addition, the influence of these variables is much stronger than for adult-directed speech. This  
16  
17 gives some explanation as to why overheard speech from adults does not significantly  
18 influence child vocabularies (Weisleder & Fernald, 2013)<sup>6</sup>. A natural assumption from this  
19 data is that greater repetitions of words and an increasing number of short words will boost  
20 children’s vocabulary growth. However, a clear caveat is the emerging effect of  
21 neighborhood density discussed earlier. Frequency and length have to give way to other  
22 variables given that with age, children’s vocabularies involve increasing amounts of  
23 infrequent and long words. A plausible factor, therefore, is neighborhood – and by providing  
24 children with a linguistic input that is littered with neighboring words (or overlapping sound  
25 sequences), their vocabulary acquisition is likely to be scaffolded based on their knowledge  
26 of neighboring words.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

44 A final note concerns the fundamental differences between child-directed speech and  
45 adult-directed speech that alter the lexical measures used in this paper. Adult-directed speech  
46 uses longer words than child-directed speech (Figure 2), and also a far greater number of  
47 word types (Table 1) that reduce word frequency (Figure 3, right panel). In all likelihood,  
48  
49  
50  
51  
52  
53

---

54 <sup>6</sup> Note there are at least three possible alternative explanations for why adult-directed speech may be  
55 less predictive of children’s vocabulary learning than child-directed speech. First, child-directed  
56 speech is more geared towards gaining the child’s attention (e.g., differences in intonation). Second, it  
57 may be the case that the language used in child-directed speech fits more with the child’s own  
58 semantic networks (e.g., see Hills, 2013). **Third, child-directed speech is produced when children are  
59 engaged in interaction, so there may be additional non-verbal cues that benefit word learning.**  
60

1  
2  
3 these will change neighborhood density, due to the presence of more word types, and  
4  
5 phonotactic probability, due to changes in word frequencies. Would our lexical measures be  
6  
7 similar across child-directed speech and adult-directed speech if the adult-directed speech  
8  
9 matched the child-directed speech on word length and number of word types? Our data  
10  
11 suggest that this may well be the case. When our lexical measures were computed from child  
12  
13 corpora (e.g., neighborhood density and phonotactic probability being computed from words  
14  
15 directed to children) there were no differences between child-directed speech and adult-  
16  
17 directed speech for neighborhood density and phonotactic probability (Table 4,  $p = .083$  and  
18  
19 Table 5,  $p = .977$  respectively).  
20  
21  
22  
23  
24

25 In summary, this paper has compared large-scale child-directed and adult-directed  
26  
27 speech to investigate whether the lexical characteristics of words in child-directed speech  
28  
29 differ from those of adult-directed speech in ways that influence the lexical properties of  
30  
31 words that appear in child productive vocabularies. We found the lexical properties of child  
32  
33 vocabularies are predicted by three of the four lexical characteristics under examination,  
34  
35 either directly or indirectly. Children learn more short than long words, more frequent than  
36  
37 infrequent words, and more neighboring words than words with few neighbors. These  
38  
39 characteristics are significantly exaggerated in child-directed speech relative to adult-directed  
40  
41 speech. Furthermore, the predictive potency of the measures change as the corpora used to  
42  
43 measure them change, with child-directed speech explaining over twice the variability in  
44  
45 child productive vocabularies than adult-directed speech. Children's word learning is clearly  
46  
47 influenced by the lexical characteristics of the words they hear, but researchers need to be  
48  
49 aware that using different corpora to calculate lexical measures may influence their perceived  
50  
51 effect on children's word learning.  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.
- Bernstein Ratner, N., & Rooney, B. (2001). How accessible is the lexicon in motherese? In J. Weissenborn, & B. Hohle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition. Vol. I*. Philadelphia: Benjamins Publishing Company.
- BNC (2007). The British National Corpus, version 3 (BNC XML Edition) [on-line database]. <http://www.natcorp.ox.ac.uk/>.
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*, B33–B44.
- Carlson, M. T., Sondregger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159–180.
- Clark, E. V., Gelman, S. A., & Lane, N. M. (1985). Compound nouns and category structure in young children. *Child Development*, *56*, 84-94.
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, *30*, 441–469.
- Dong, S., Gu, Y., & Vigliocco, G. (2021). The impact of child-directed language on children's lexical development. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*, 1444–1450.
- Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.
- Forrester, M. (2002). Appropriating cultural conceptions of childhood: Participation in conversation. *Childhood*, *9*, 255-278.

## Child- and adult-directed speech

37

- 1  
2  
3 French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: a recognition-based  
4  
5 connectionist framework for sequence segmentation and chunk extraction.  
6  
7 *Psychological Review*, *118*, 614–636.  
8  
9
- 10 Furrow, D., Nelson, K., & Benedict, H. (1979). Mothers' speech to children and syntactic  
11  
12 development: Some simple relationships. *Journal of Child Language*, *6*, 423–442.  
13
- 14 Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in  
15  
16 the development of vocabulary in children: A longitudinal study. *Journal of Memory  
17  
18 and Language*, *28*, 200–213.  
19
- 20 Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to me:  
21  
22 The social context of infant-directed speech and its effects on early language  
23  
24 acquisition. *Current Directions in Psychological Science*, *24*, 339–344.  
25  
26
- 27 Hartig, F. (2022). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)  
28  
29 Regression Models*. R package version 0.4.5. [https://CRAN.R-  
30  
31 project.org/package=DHARMA](https://CRAN.R-project.org/package=DHARMA)  
32  
33
- 34 Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special case  
35  
36 of motherese? *Journal of Child Language*, *15*, 395–410.  
37  
38
- 39 Henry, A. (1995). *Belfast English and standard English: Dialect variation and parameter  
40  
41 setting*. New York: Oxford University Press.  
42  
43
- 44 Hills, T. (2013). The company that words keep: Comparing the statistical structure of child-  
45  
46 versus adult-directed language. *Journal of Child Language*, *40*, 586–604.  
47  
48
- 49 Hoff, E., & Naigles, L. R. (2002). How children use input to acquire a lexicon. *Child  
50  
51 Development*, *73*, 418–433.  
52  
53
- 54 Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary  
55  
56 growth: Relation to language input and gender. *Developmental Psychology*, *27*, 236–  
57  
58 248.  
59  
60

- 1  
2  
3 Jones, G., Cabiddu, F., Andrews, M., & Rowland, C. (2021). Chunks of phonological  
4 knowledge play a significant role in children's word learning and explain effects of  
5 neighborhood size, phonotactic probability, word frequency and word length. *Journal*  
6 *of Memory and Language, 119*, 104232.  
7  
8  
9  
10  
11  
12 Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using  
13 computational modeling to isolate the effects of the quantity and the diversity of the  
14 input on vocabulary growth. *Cognitive Psychology, 98*, 1–21.  
15  
16  
17  
18  
19 Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and  
20 frequency in specific language impairment: Accuracy and error data from two  
21 nonword repetition tests. *Journal of Speech Language and Hearing Research, 53*,  
22 1642–1655.  
23  
24  
25  
26  
27  
28 Jones, G., & Witherstone, H. L. (2011). Lexical and sublexical knowledge influences the  
29 encoding, storage, and articulation of nonwords. *Memory & Cognition, 39*, 588–599.  
30  
31  
32  
33 Lieven, E., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of  
34 multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20*, 481–507.  
35  
36  
37  
38 Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech  
39 clarity and infants' speech discrimination skills. *Developmental Science, 6*, F1–F10.  
40  
41  
42  
43 MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



- 1  
2  
3 McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed  
4  
5 speech and the development of speech perception: Enhancing development or an  
6  
7 unintended consequence? *Cognition*, *129*, 362–378.  
8  
9  
10 Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early  
11  
12 word learning environments. *Cognitive Science*, *42*, 375–412.  
13  
14 Naigles, L. R., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs?  
15  
16 Effects of input frequency and structure on children's early verb use. *Journal of Child*  
17  
18 *Language*, *25*, 95–120.  
19  
20  
21 Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of*  
22  
23 *Memory and Language*, *39*, 246–263.  
24  
25  
26 R Core Team (2018). *R: A language and environment for statistical computing*. R  
27  
28 *Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>.  
29  
30  
31 Rodriguez-Cuadrado, S., Baus, C., & Costa, A. (2018). Foreigner talk through word  
32  
33 reduction in native/non-native spoken interactions. *Bilingualism*, *21*, 419–426.  
34  
35  
36 Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-  
37  
38 directed speech vocabulary development. *Child Development*, *83*, 1762–1774.  
39  
40  
41 Rowland, C. F. & Fletcher, S. L. (2006). The effect of sampling on estimates of lexical  
42  
43 specificity and error rates. *Journal of Child Language*, *33*, 859-877.  
44  
45  
46 Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density  
47  
48 longitudinal corpus. In *Proceedings of the 31st Annual Conference of the Cognitive*  
49  
50 *Science Society*. Amsterdam, Netherlands.  
51  
52  
53 Saxton, M. (2009). The inevitability of child directed speech. In S. Foster-Cohen (Ed.)  
54  
55 *Language Acquisition*, pp. 62–86.  
56  
57  
58 Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in  
59  
60 toddlers. *Journal of Speech, Language, and Hearing Research*, *53*, 670–683.

## Child- and adult-directed speech

- 1  
2  
3 Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and  
4 semantic variables on word learning by infants. *Journal of Child Language*, *36*, 291–  
5 321.  
6  
7  
8  
9  
10 Storkel, H. L., & Lee, S. (2011). The independent effects of phonotactic probability and  
11 neighborhood density on lexical acquisition by preschool children. *Language and*  
12 *Cognitive Processes*, *26*, 191–211.  
13  
14  
15  
16  
17 Suttora, C., Salerni, N., Zanchi, P., Zampini, L., Spinelli, M., & Fasolo, M. (2017).  
18 Relationships between structural and acoustic properties of maternal talk and  
19 children's early word recognition. *First Language*, *37*, 612–629.  
20  
21  
22  
23  
24 Swingle, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning  
25 of specific English words. *Child Development*, *89*, 1247–1267.  
26  
27  
28  
29 Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of  
30 performance limitations in the acquisition of verb-argument structure: An alternative  
31 account. *Journal of Child Language*, *28*, 127–152.  
32  
33  
34  
35  
36 Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word  
37 segmentation. *Infancy*, *7*, 53–71.  
38  
39  
40 Tommerdahl, J., & Kilpatrick, C. (2013). Analyzing reliability of grammatical production in  
41 spontaneous samples of varying length. *Journal of Child Language Teaching and*  
42 *Therapy*, *29*, 171-183.  
43  
44  
45  
46  
47 Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine phonotactic online dictionary,  
48 version 2. <http://www.iphod.com>.  
49  
50  
51  
52 Walley, A. C. (1993). The role of vocabulary development in children's spoken word  
53 recognition and segmentation ability. *Developmental Review*, *13*, 286–350.  
54  
55  
56  
57  
58  
59  
60

Child- and adult-directed speech

41

1  
2  
3 Walley, A. C., Metsala, J. L., & Victoria, M. (2003). Spoken vocabulary growth: Its role in  
4 the development of phoneme awareness and early reading ability. *Reading and*  
5  
6

7 *Writing: An Interdisciplinary Journal*, 16, 5–20.

8  
9  
10 Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
strengthens processing and builds vocabulary. *Psychological Science*, 24, 2143–2152.

Wells, C. G. (1981). *Learning through interaction: The study of language*

*development*. Cambridge, UK: Cambridge University Press.

You, G., Bickel, B., Daum, M. M., & Stoll, S. (2021). Child-directed speech is optimized for  
syntax-free semantic inference. *Scientific Reports*, 11, 1–11.

## Appendix 1. Plots and analyses when excluding all compound nouns.

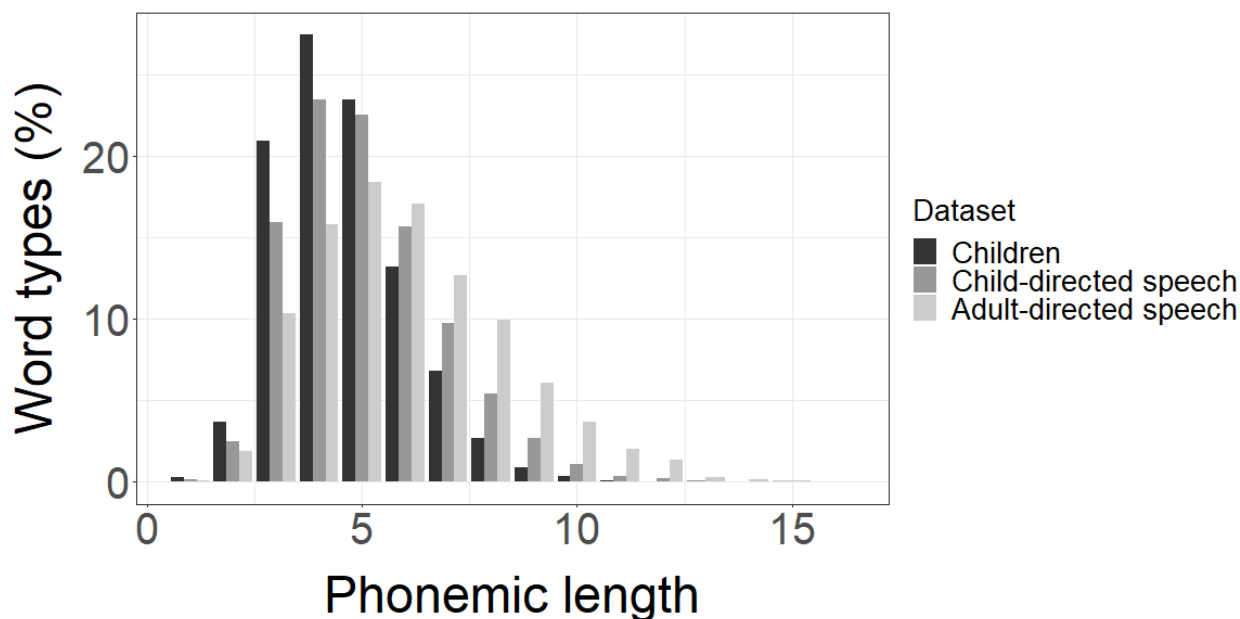


Figure A1.1. Percentage of word types varying in phonemic length, for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech) when excluding all compound nouns.

Table A1.1. Kolmogorov-Smirnov comparisons for phonemic word length between different corpora when excluding all compound nouns, with 95% bootstrapped confidence intervals (1,000 iterations). *p* values and confidence intervals were corrected using Holm's correction.

Comparison	D	p	Lower Bci	Upper Bci
Children vs. Child-directed speech	.11	< .001	.09	.13
Child- vs. Adult-directed speech	.18	< .001	.16	.2
Children vs. Adult-directed speech	.29	< .001	.27	.32

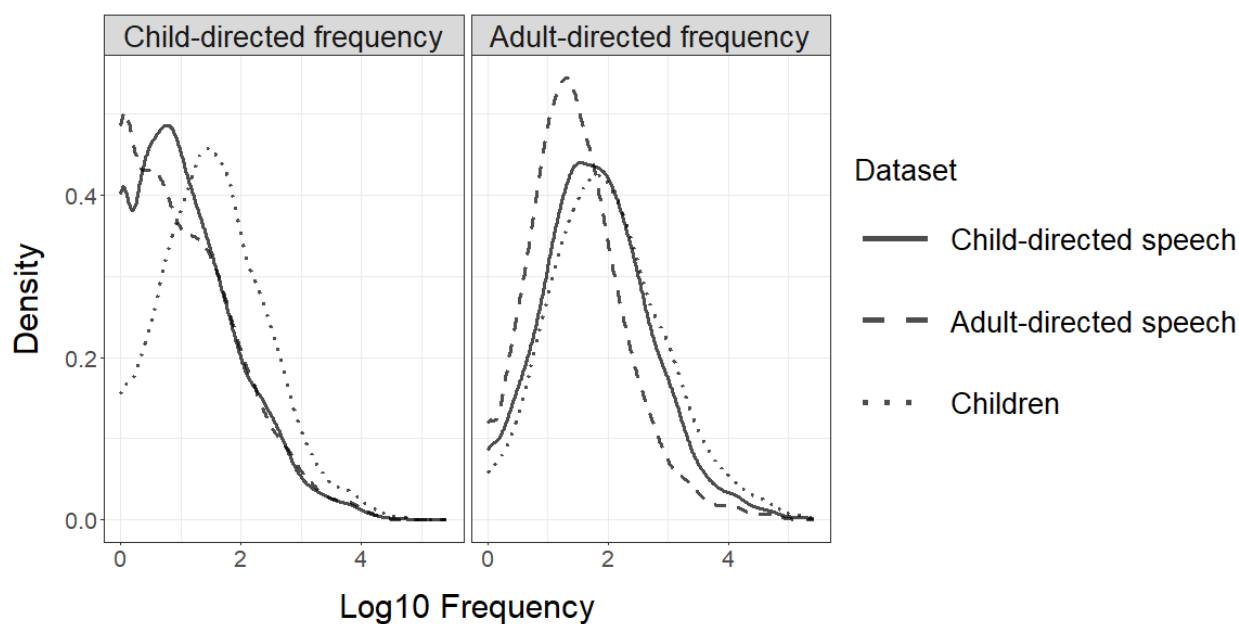


Figure A1.2. Density plot of word types (excluding compound nouns) varying in Log10 word frequency, for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech) and with word frequency computed using child-directed speech (left panel) and adult-directed speech (right panel).

Table A1.2. Kolmogorov-Smirnov comparisons for Log10 word frequency between different corpora (excluding compound nouns), with 95% bootstrapped confidence intervals (1,000 iterations). Comparisons are performed when computing Log10 word frequency using child-directed speech and adult-directed speech.

Comparison	Measure	D	p	Lower Bci	Upper Bci
Children vs. Child-directed speech	Child-directed	.25	< .001	.22	.27
Child- vs. Adult-directed speech	Child-directed	.06	< .001	.04	.09
Children vs. Adult-directed speech	Child-directed	.26	< .001	.23	.3
Children vs. Child-directed speech	Adult-directed	.07	< .001	.06	.1
Child- vs. Adult-directed speech	Adult-directed	.18	< .001	.16	.21
Children vs. Adult-directed speech	Adult-directed	.25	< .001	.22	.28

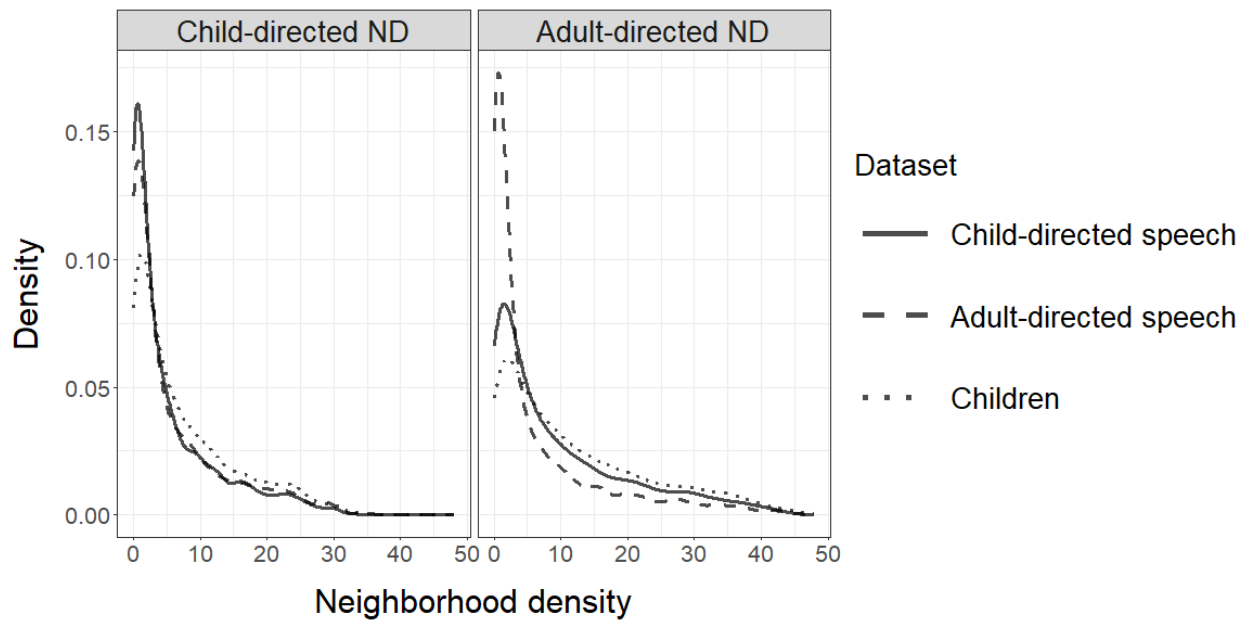


Figure A1.3. Density plot of word types varying in neighborhood density (excluding compound nouns), for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech) and with neighborhood density computed using child-directed speech (left panel) and adult-directed speech (right panel).

Table A1.3. Kolmogorov-Smirnov comparisons for neighborhood density between different corpora (excluding compound nouns), with 95% bootstrapped confidence intervals (1,000 iterations). Comparisons are performed when computing neighborhood density using child-directed speech and adult-directed speech.

Comparison	Measure	D	p	Lower Bci	Upper Bci
Children vs. Child-directed speech	Child-directed	.14	< .001	.11	.16
Child- vs. Adult-directed speech	Child-directed	.03	.455	.02	.05
Children vs. Adult-directed speech	Child-directed	.14	< .001	.1	.17
Children vs. Child-directed speech	Adult-directed	.1	< .001	.08	.13
Child- vs. Adult-directed speech	Adult-directed	.2	< .001	.18	.23
Children vs. Adult-directed speech	Adult-directed	.3	< .001	.27	.33

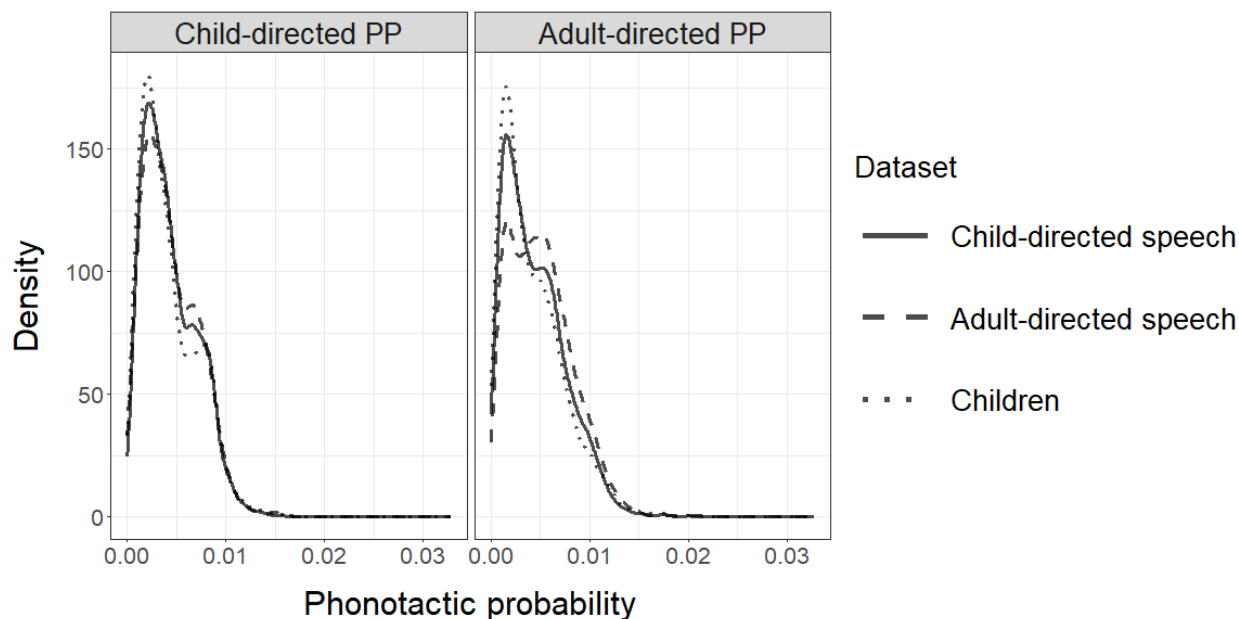


Figure A1.4. Density plot of word types (excluding compound nouns) varying in phonotactic probability, for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech) and with phonotactic probability computed using child-directed speech (left panel) and adult-directed speech (right panel).

Table A1.4. Kolmogorov-Smirnov comparisons for phonotactic probability between different corpora (excluding compound nouns), with 95% bootstrapped confidence intervals (1,000 iterations). Comparisons are performed when computing phonotactic probability using child-directed speech and adult-directed speech.

Comparison	Measure	D	p	Lower Bci	Upper Bci
Children vs. Child-directed speech	Child-directed	.05	.003	.03	.08
Child- vs. Adult-directed speech	Child-directed	.02	.614	.02	.06
Children vs. Adult-directed speech	Child-directed	.06	.001	.03	.1
Children vs. Child-directed speech	Adult-directed	.05	< .001	.03	.08
Child- vs. Adult-directed speech	Adult-directed	.1	< .001	.08	.13
Children vs. Adult-directed speech	Adult-directed	.15	< .001	.12	.18

Appendix 2. Plots for individual word categories (noun, verb, adjective, and 'other').

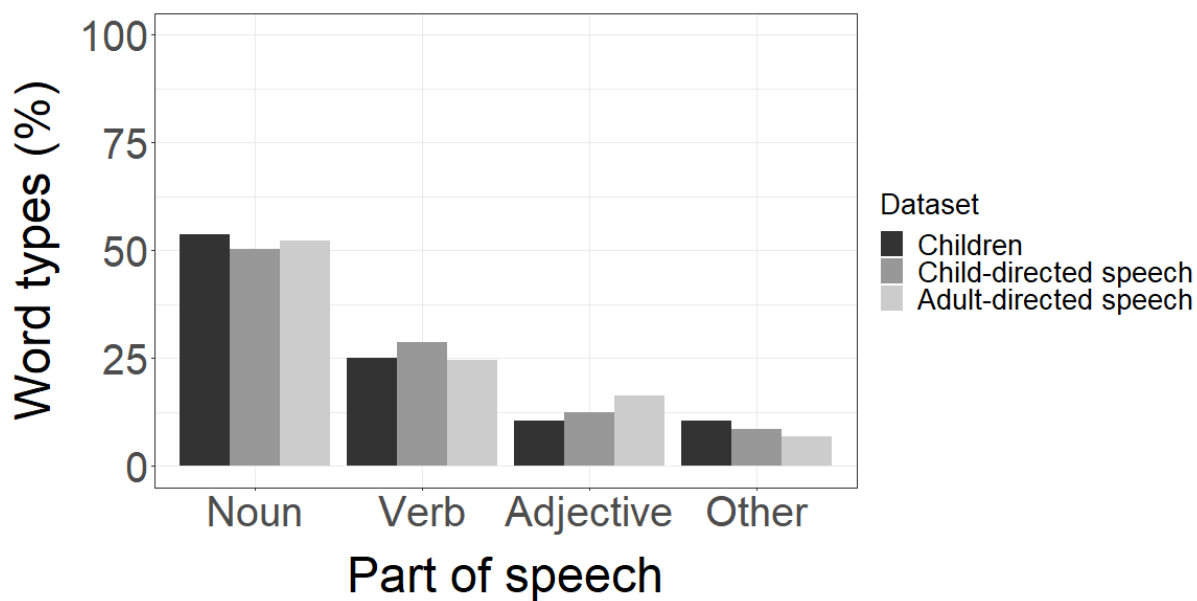


Figure A2.1. Distributions of each word category within the different corpora (children, child-directed speech, adult-directed speech).

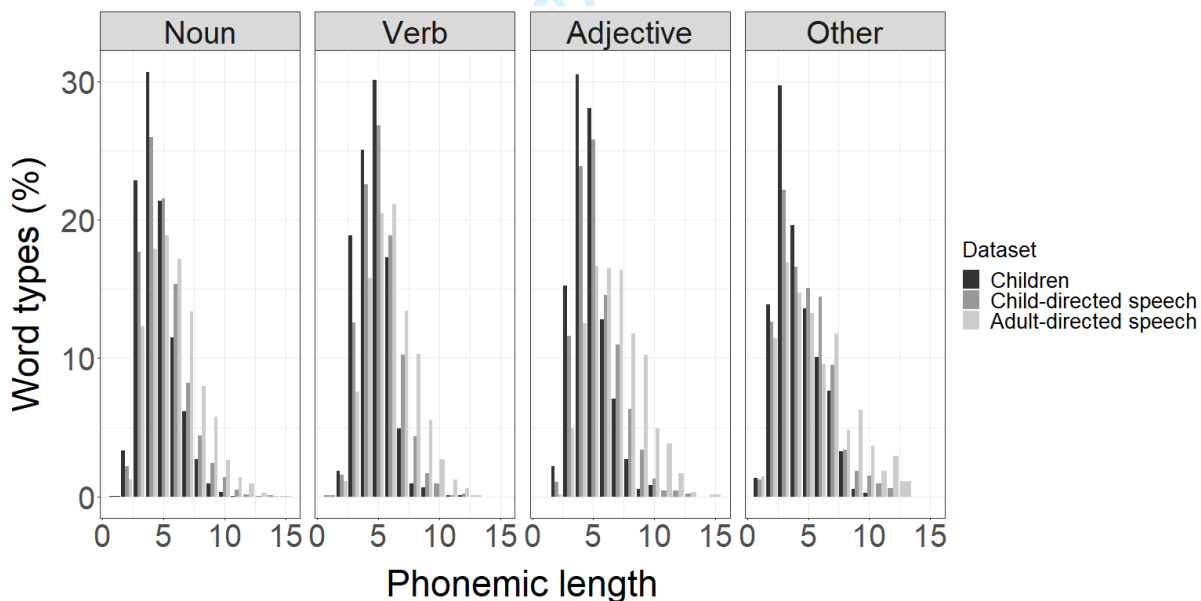


Figure A2.2. Distributions of phonemic length across each word category within the different corpora (children, child-directed speech, adult-directed speech).



Child- and adult-directed speech

47

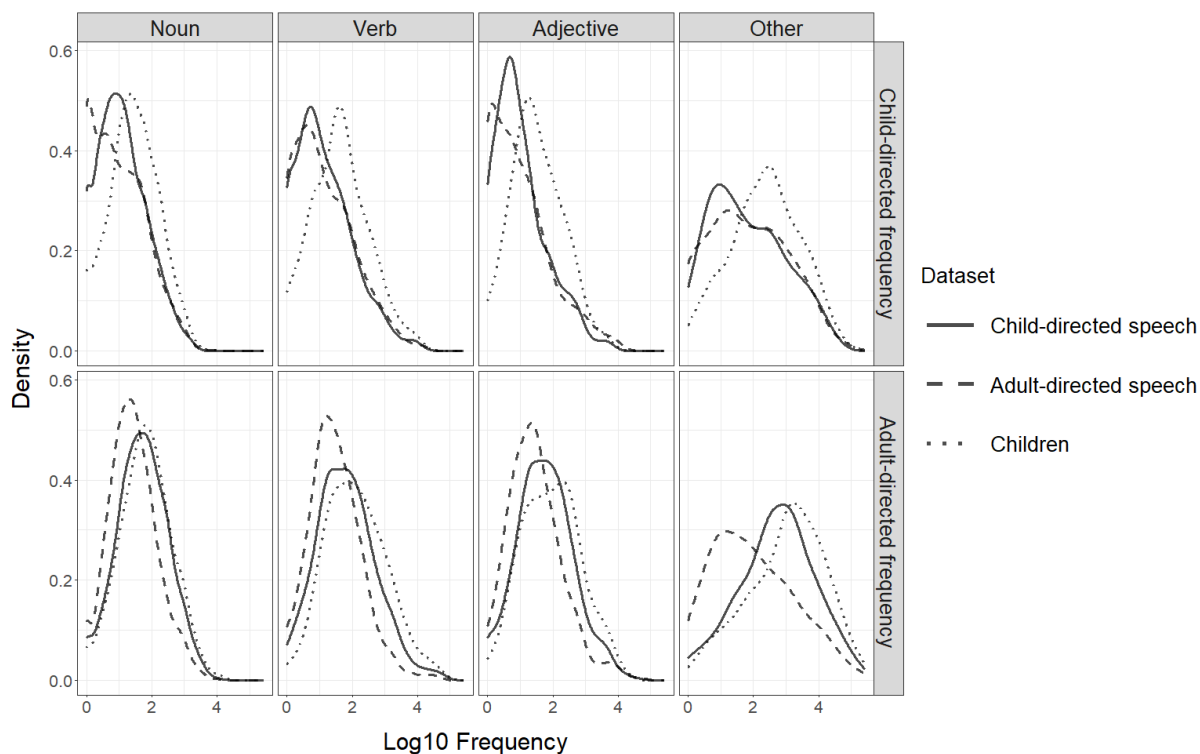
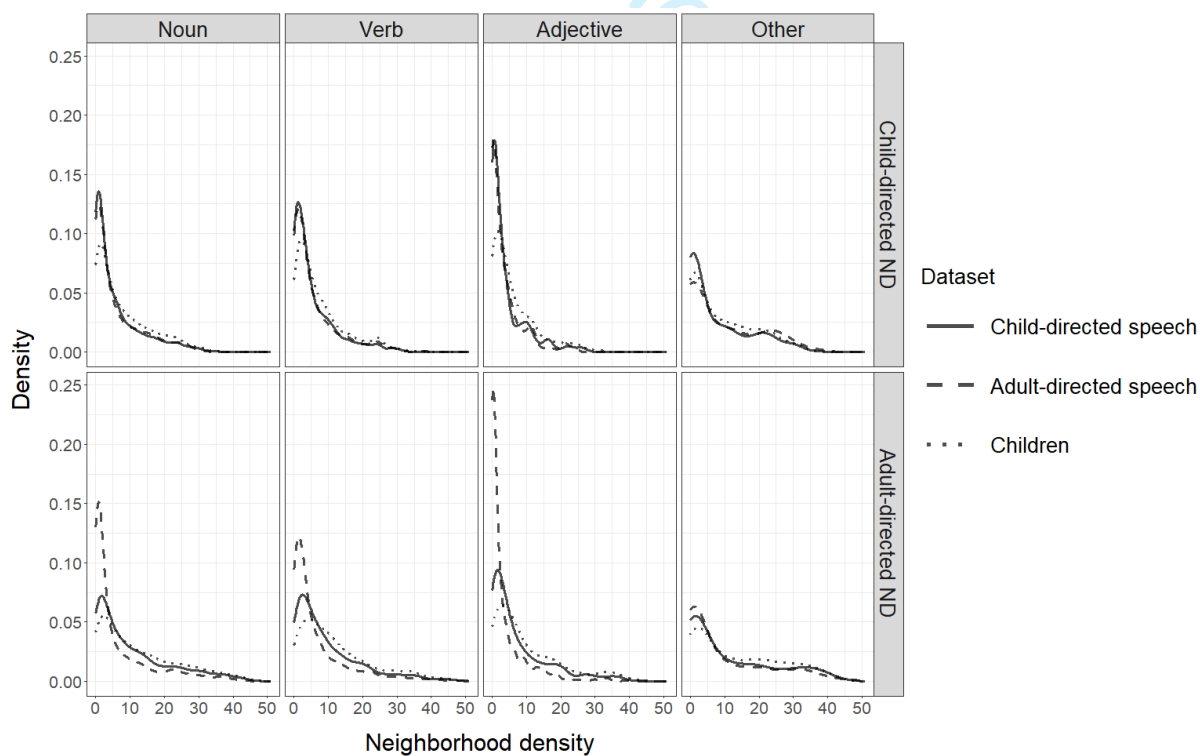


Figure A2.3. Density plot of word types varying in Log10 word frequency, for each word category within each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech).



## Child- and adult-directed speech

48

Figure A2.4. Density plot of word types varying in neighborhood density, for each word category and separated by sample corpus (child productive vocabularies, child-directed speech, adult-directed speech).

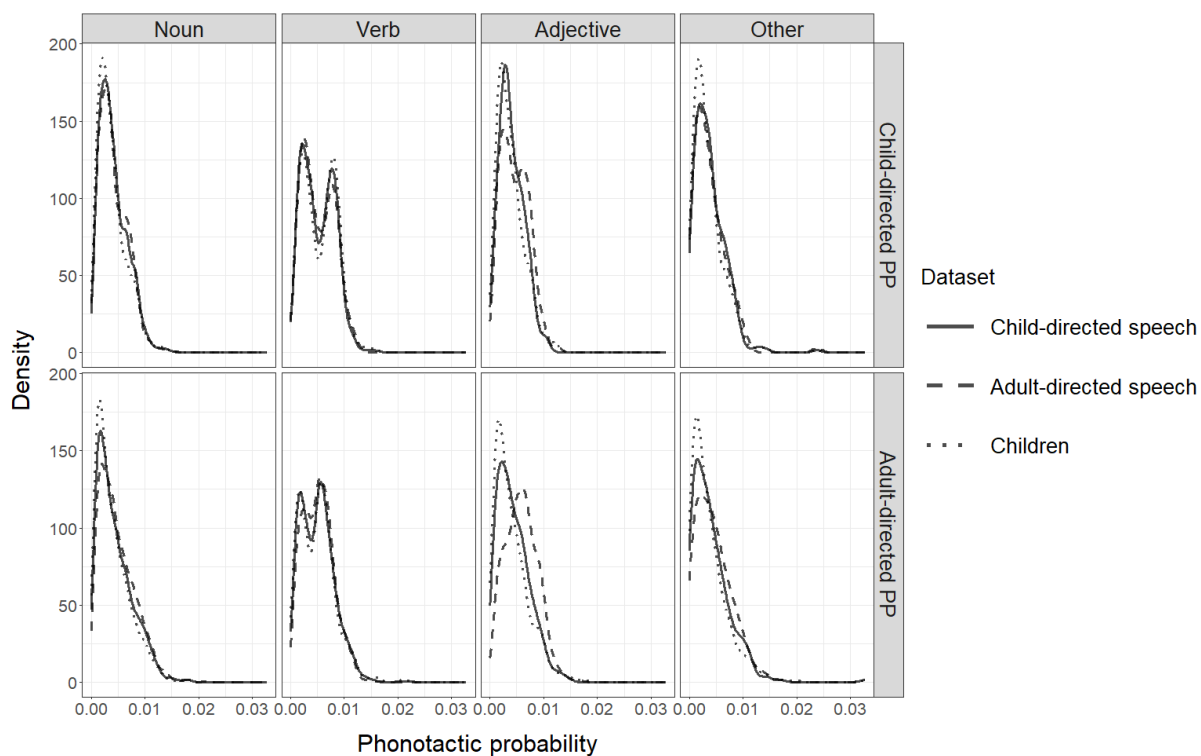


Figure A2.5. Density plot of word types varying in phonotactic probability, for each word category and for each sample corpus (child productive vocabularies, child-directed speech, adult-directed speech).