



### **The Most Frequently Used Words: Comparing Child-directed Speech and Young Children's Speech to Inform Vocabulary Selection for Aided Input**

Journal:	<i>Augmentative and Alternative Communication</i>
Manuscript ID	TAAC-2018-0046.R2
Manuscript Type:	Research Article
Keywords:	Core Vocabulary, AAC, Vocabulary Selection, Maternal Input

SCHOLARONE™  
Manuscripts

Use of Core Vocabulary in Maternal Input with Emerging Communicators

Nancy Quick, Karen Erickson, and Jacob Mccright

University of North Carolina at Chapel Hill

Author Note

Nancy Quick and Karen Erickson, The Center for Literacy and Disability Studies, University of North Carolina at Chapel Hill; Jacob Mccright, University of North Carolina at Chapel Hill

This work was supported by U.S. Department of Education, Office of Special Education Programs Grant No. H327S160005. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service, or enterprise mentioned in this publication is intended or should be inferred.

Correspondence concerning this article should be addressed to Nancy Quick, The Center for Literacy and Disability Studies, University of North Carolina at Chapel Hill, 321 S. Columbia St, Suite 1100, Chapel Hill, NC 27599-7335. Email: [nancy\\_quick@med.unc.edu](mailto:nancy_quick@med.unc.edu)

### Abstract

Transactional theories of communication development focus on the interplay among child, caregiver, and environmental variables. Typically, this interplay involves symmetry between receptive and expressive modes (i.e., speech), but is asymmetrical for children with complex communication needs who hear speech but use graphic symbols expressively. Aided input, during which a communication partner points to graphic symbols while talking, may increase symmetry, but it is challenging to determine which words to represent with graphic symbols to ensure adequate aided input is provided. In this study, secondary analysis of transcripts of 16 mothers who interacted with their children with typical development across 6 time points (between 9 and 15 months) revealed 267 words that comprised 80% of the 257,480 words the mothers used. This list of words that mothers used most frequently was compared to three existing lists of the expressive vocabulary used most frequently by 65 toddlers and preschoolers with typical development, indicating substantial overlap. The results suggest that there is a common set of frequently occurring words that mothers use in their daily interactions with infants and toddlers, and that these same words also comprise a significant proportion of the words most frequently used by young children. Implications for representing these frequently occurring words with graphic symbols on the communication systems of children with complex communication needs are discussed.

*Keywords:* Core Vocabulary; Augmentative and alternative communication; Vocabulary selection; Maternal input

The Most Frequently Used Words: Comparing Child-directed Speech and Young Children's  
Speech to Inform Vocabulary Selection for Aided Input

Child language outcomes are influenced by the language children hear around them every day. The sheer number of words infants and toddlers hear their caregivers use is associated with their rate of vocabulary development and speed of later vocabulary processing (Hurtado, Marchman, & Fernald, 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). Transactional theories of language development focus on the interplay between child and caregiver, as well as the environment (McLean & McLean, 1999). Typically, the interplay involves symmetry between receptive and expressive modes (i.e., speech) but is asymmetrical for children with complex communication needs who hear speech but are learning to use graphic symbols expressively. Research suggests that aided input may offer a means of reducing this asymmetry between input and output modes for these children (O'Neil, Light, & Pope, 2018).

Aided input, during which a communication partner points to graphic symbols while talking (von Tetzchner, 2015), supports children in learning and using graphic symbols to communicate (O'Neil et al., 2018; Sennott, Light, & McNaughton, 2016). Unfortunately, young children with complex communication needs rarely see others communicate with the types of graphic symbols they use (Ballin, Balandin, Stancliffe, & Togher, 2011; von Tetzchner, 2015), even when the symbols are readily available to caregivers (Trudeau, Cleave, & Woelk, 2003). Furthermore, there are few supports to guide caregivers in selecting which words to represent with graphic symbols. Given that studies focused on diversity and overall quantity of vocabulary in caregiver input consistently point to a relationship between this input and child outcomes (Hoff, 2003; Pan, Rowe, Singer, & Snow, 2005; Rowe, 2012), evidence-based decisions

1  
2  
3 regarding the type of words to represent with graphic symbols to increase input-out symmetry  
4  
5 are critical.

6  
7  
8 One of the many challenges caregivers face in using aided input to support a child's  
9  
10 receptive and expressive language development is having access to graphic symbols that  
11  
12 represent words that are meaningful in each communication exchange. Selecting graphic  
13  
14 symbols that are context specific works to support some interactions (Drager, Light, Speltz,  
15  
16 Fallon, & Jeffries, 2003; Light et al., 2004) but doesn't meet moment-to-moment needs and  
17  
18 likely narrows the child's generative understanding and use of the symbols once they are learned  
19  
20 (von Tezchner, 2015). Selecting graphic symbols that can be used in a variety of contexts  
21  
22 increases the likelihood that a child will learn them because language learning in general is  
23  
24 dependent upon frequent input and output opportunities with varied exemplars across contexts  
25  
26 (Poll, 2011; Tomasello, 2003). Words that are used most frequently in natural speech are  
27  
28 employed for a variety of purposes across contexts.  
29  
30  
31

32  
33 Words that are used most frequently across contexts in oral and written language  
34  
35 comprise core vocabulary (Banajee, Dicarolo, & Stricklin, 2003; Deckers, Van Zaalen, Van  
36  
37 Balkom, & Verhoeven, 2017; Trembath, Balandin, & Togher, 2007). Rather than specific nouns,  
38  
39 core vocabulary is composed primarily of pronouns, verbs, auxiliary verbs, prepositions,  
40  
41 adjectives, and determiners. The existence of words that can be characterized as core vocabulary  
42  
43 has been confirmed through a number of studies investigating young children with typical  
44  
45 development (e.g., Banajee et al., 2003), mono and bilingual school-aged children with and  
46  
47 without language impairments (e.g., Boenisch & Soto, 2015), children with intellectual  
48  
49 disabilities (Deckers, et al., 2017), and adults (e.g., Balandin & Iacono, 1999). For children who  
50  
51 are learning to use graphic symbols to communicate, these lists often guide vocabulary selection  
52  
53  
54  
55  
56  
57  
58  
59

1  
2  
3 because the words are used so widely that they offer robust flexibility in communicating for a  
4  
5 broad range of purposes across a variety of contexts (Dennis, Erickson, & Hatch, 2013; Geist,  
6  
7 Hatch & Erickson, 2014; Van Tatenhove, 2009).  
8  
9

10           Graphic symbols are intended to stand for or represent a thing or concept (Alant,  
11  
12 Bornman, & Lloyd, 2006). These representations vary in form and can include photographs of  
13  
14 objects or people; colored or black-and-white line drawings; or more abstract symbols,  
15  
16 lexigrams, or printed words (Beukelman & Mirenda, 2015). Graphic symbols also vary with  
17  
18 respect to their iconicity, or the degree to which they represent their referents (Lloyd & Fuller,  
19  
20 1990). Symbols that represent core vocabulary are not iconic because core words are  
21  
22 conceptually referenced and do not include nouns or other word-types that are easily represented  
23  
24 with graphic symbols (Snodgrass, Stoner, & Angell, 2013). As a result of this limited iconicity,  
25  
26 core words are not often selected as the focus of an initial lexicon for children who must learn to  
27  
28 use graphic symbols to communicate (Schlosser, 2003).  
29  
30  
31  
32

33           It is the case, that, in the absence of instruction or experience, iconicity influences a  
34  
35 child's ability to understand and use a graphic symbol (Simcock & DeLoache, 2006); however,  
36  
37 general experience with pictures, understanding of the words that symbols represent,  
38  
39 understanding of the intent of the person using the symbol, and specific instruction in how to use  
40  
41 the symbol, have a greater impact and can mediate success with graphic symbols that lack  
42  
43 iconicity (Stephenson, 2009). In addition, iconicity has less influence on very young children  
44  
45 who are learning the meaning and use of graphic symbols than it does on older children who  
46  
47 have an understanding of symbolic representation (Namy, 2001; Namy, Campbell, & Tomasello,  
48  
49 2004). In other words, there is reason to believe that providing parents with a set of graphic  
50  
51 symbols that represent the words they use most often in their child-directed speech, and  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 supporting them in providing aided input with those symbols, would provide their very young  
4  
5 children with the conditions they need to successfully learn and use those symbols, despite the  
6  
7 limited iconicity of the symbols themselves.  
8  
9

10 One purpose of existing studies of core vocabulary has been to identify vocabulary that  
11  
12 can be represented by graphic symbols for expressive use by individuals with complex  
13  
14 communication needs. For children who use AAC, it has been proposed that productive  
15  
16 vocabulary studies of young children (e.g., Banajee, et al., 2003) inform vocabulary selection for  
17  
18 their expressive use (Trembath, et al., 2007). However, no studies have considered the words that  
19  
20 children are most likely to hear in their interactions with caregivers beginning in infancy in order  
21  
22 to inform their receptive use during aided input. Identifying the words children hear most often  
23  
24 could improve input-output symmetry by maximizing opportunities to use aided input to  
25  
26 demonstrate communication with the same mode that children with complex communication  
27  
28 needs are expected to use expressively. Input-output symmetry could be improved further by  
29  
30 identifying the subset of the most frequently used words of caregivers that overlap with the  
31  
32 words most frequently used by young children identified in previous studies. This would allow  
33  
34 children with complex communication needs to learn the meaning and use of the same graphic  
35  
36 symbols they are most likely to use in their expressive communication.  
37  
38  
39  
40  
41

42 The purpose of the current study was to determine which words mothers use most  
43  
44 frequently in speech they direct to their infants and toddlers. A secondary purpose was to  
45  
46 compare the maternal lists with lists of words used most frequently by young children who are  
47  
48 developing typically. The analyses were designed to determine how many different words  
49  
50 mothers used, which words they used most frequently, and how the words they used most  
51  
52 frequently compared with words young children used most frequently.  
53  
54  
55  
56  
57  
58  
59  
60

## Method

This study employed secondary data analyses of several existing data sets. The first was the maternal vocabulary in transcriptions of mother-child interactions extracted from the Brent corpus (Brent & Siskind, 2001) of the CHILDES (MacWhinney, 2000) database. Available demographic information about the Brent corpus is reported below. The other data were the words lists reported by Banajee and colleagues (2003); Beukelman, Jones, and Rowan (1989); and Marvin, Beukelman, and Bilyeu (1994) in their studies of words that older toddlers and preschoolers used most frequently in their expressive communication across a variety of contexts as described in the sections that follow.

### Maternal Vocabulary

For the Brent corpus, 16 mothers were recruited through advertisements in a free newspaper distributed in and around Baltimore Maryland. The education levels were one master's degree, six bachelor's degrees, six some years of college, two high school degrees, and one unreported. The mother-child dyads were recorded at home approximately every 2 weeks when the children were between 9 and 15 months of age. During the recordings, mothers were asked to maintain their normal routines, with the exception of avoiding activities that included media (e.g., TV, radio) or long telephone conversations. Each session was 90-120 min in length, and the middle 75 min were originally transcribed and coded using Computer Language Analysis (CLAN; MacWhinney, 2000). The Brent corpus included a total of 179 samples, and the number of samples for each mother-child dyad ranged from six to 14. For the current investigation, the first six samples of each dyad were used, for a total of 96 samples.



## Child Vocabulary

Three existing word lists were used to compare the words mothers used most frequently with the words children used most frequently. The first list was generated by Banajee and colleagues (2003) from the language samples (150 utterances each) of 50 toddlers (34 were girls) with typical development aged 24 to 36 months. The children had been recruited from five different preschools, and language samples were collected during two different activities (play and snack). The second list was generated by Marvin and colleagues (1994) from the language samples of 10 preschool-aged children with typical development (range: 48 to 62 months), including three females and seven males. The samples for Marvin et al. had been collected during 2- to 2.5-hr of naturally occurring interactions at home and school. The third list was generated by Beukelman and colleagues (1989) from 3000-word language samples collected from six pre-school-aged children with typical development (range: 44 to 57 months) as they engaged in typical interactions at school.

## Procedures

**Cleaning and coding maternal transcripts.** The original transcripts of mother-child interactions that served as the source material for Brent and Siskend (2001) were downloaded into spreadsheets. These spreadsheets were then processed using scripts in the data analysis package Pandas (McKinney, 2010), a module found within the Python 3.6 (Python Software Foundation, 2016). Pandas supports text-based analysis of large data sets and supports the identification and coding of lexical items that were spoken, sung, or cited so they could be grouped together for separate analysis. Spoken lexical items were stripped of extraneous punctuation, standardized as lower-case, and broken apart using standard white space based tokenization.

1  
2  
3       **Identifying real words.** The composite sample of 267,204 tokens (i.e., verbalized words  
4 and nonwords) was reviewed by two researchers to exclude the following: utterances of other  
5 adults ( $n = 1,373$  tokens), unintelligible utterances ( $n = 2,575$  tokens), letters (e.g., abc;  $n = 343$ ),  
6 and proper names of children or adults in the sample (e.g., Emily;  $n = 3,749$  tokens). Then a two-  
7 step process was completed to determine whether the remaining tokens would be coded as words  
8 and included in further analysis. This process first involved checking each token in the  
9 Educators' Word Frequency Guide, which provides an indicator of both the frequency and  
10 dispersion of words in written English in texts at the primary and secondary levels  
11 (Zeno, Ivens, Millard, & Duvvuri, 1995). Tokens that did not appear in this resource were  
12 manually entered into the Merriam-Webster online dictionary ([https://www.merriam-  
14 webster.com/](https://www.merriam-<br/>13 webster.com/)). Those tokens that could not be found in either resource were coded as nonwords  
15 (e.g., "mooshieboos";  $n = 1,684$  tokens) and removed from further analysis. Tokens that were  
16 found in one or both resources were coded as words.

17  
18       **Data analysis.** Frequency counts were generated for each word in the maternal  
19 transcripts based on its occurrence across all session transcripts. Relative frequency was  
20 calculated for each word by dividing the frequency count of each word by the total number of  
21 words in the composite sample. After rank ordering the words on the basis of their relative  
22 frequency, cumulative frequency was calculated by adding the relative frequencies of a given  
23 word and all words with a greater relative frequency. A commonality score of 1 to 16 was  
24 assigned to each word, reflecting the number of mothers who produced the word one or more  
25 times. After the data were analyzed to determine the most frequently occurring words with the  
26 highest commonality scores in the maternal transcripts, each word was associated with a binary  
27 coding (1 for success, 0 for failure) to describe whether the word was also found on one or more  
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

of the child word frequency lists that were identified in the three studies of expressive word frequency among young children (Banajee et al., 2003; Beukelman et al., 1989; Marvin et al., 1994).

## Results

### Analysis of Maternal Input

The secondary data analysis began with an examination of the maternal input of the 16 mothers drawn from the Brent corpus (Brent & Siskend, 2001). For each session, an analysis of the language transcript was completed with CLAN, using a variety of word and sentential measures. For all 16 mothers, the average of each measure (i.e., mean number of utterances, mean length of utterance – words, mean length of utterance – morphemes, number of total words, number of different words, type-token ratios) was computed across six sessions (see Table 1). There was a great deal of variety in the amount of maternal input. During the 75-min transcripts, the total number of utterances ranged from approximately 250 to 1400. While there was also a wide range in the total number of words used in each session (approximately 850 to 5,600), across the sessions, there was a substantially narrower range in the total number of different words (nearly 200 to 600).

Insert Table 1 about here

The process to identify the real words among the tokens resulted in a total of 257,480 words, including 5,187 different words. Using the criteria outlined by Trembath and colleagues (2007), core vocabulary was defined as words in the composite sample that occurred with a relative frequency of at least 0.5 per 1,000 and were used by at least half of the participants (i.e.,  $n \geq 8$ ). These criteria resulted in a list of the 267 most frequently occurring words (5.1% of the

number of the different words). The number of words that only occurred once composed one-third of the number of different words in the composite sample ( $n = 1,730$ ; 33.35%).

Importantly, the relatively small set of core vocabulary words composed most of the total words in the composite sample. The multiple uses of the 267 most frequently occurring words represented 80.15% of all of the words (See Figure 1). Furthermore, approximately one-third of the composite language sample is represented by the 25 most frequently occurring words, nearly one-half is represented by the 50 most frequently occurring words, and two-thirds is represented by the 125 most frequently occurring words (See Table 2).

Insert Table 2 and Figure 1 about here

A list of the most frequently occurring words, including their relative frequency of occurrence and commonality score, is reported in Appendix A (online Supplemental only). As evidenced by the commonality scores, the most frequently occurring words were also used by most of the mothers. Approximately 65% ( $n = 173$ ) of the most frequently occurring words were produced by all 16 of the mothers in the sample, and 93% ( $n = 248$ ) were produced by 13 or more mothers.

### Comparing Maternal and Child Word Lists

The second part of the secondary data analysis compared the most frequently occurring words from the 16 mothers drawn from the Brent corpus (Brent & Siskend, 2001) to the most frequently occurring words of young children drawn from each of three studies (Banajee et al., 2003; Beukelman et al., 1989; Marvin et al., 1994). There was considerable overlap between the most frequently occurring words produced in the infant-directed speech of mothers and previous word frequency studies of young children's expressive vocabulary use. For example, the 23 words used most frequently by toddlers in the study reported by Banajee and colleagues

1  
2  
3 appeared in the composite sample from the mothers, with more than 90% overlapping with the  
4 vocabulary the mothers used most frequently. The only words that appeared in the most  
5 frequently occurring set of words identified by Banajee and colleagues that were not used  
6 frequently by the mothers from the Brent corpus were “finished” and “mine.” Similarly, 238 of  
7 the 250 core words used most frequently by preschoolers in the study by Beukelman and  
8 colleagues appeared in the composite sample from the mothers, and nearly two-thirds of the  
9 words in Beukelman et al. overlapped with the words mothers used most frequently. Over half of  
10 the discrepancies between these two word lists were due to the inclusion of participants’ names  
11 in the study by Beukelman and colleagues. Finally, 324 of the 332 words used by preschoolers in  
12 the study by Marvin and colleagues appeared in the composite sample of maternal vocabulary  
13 from the Brent corpus, with a little more than one half of those words overlapping with the most  
14 frequently occurring words produced by mothers. Direct comparisons of the most frequently  
15 occurring words across the four studies are illustrated in Table 3.

16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33 Insert Table 3 about here

### 34 35 **Discussion**

36  
37 This study sought to identify and describe the words used by mothers when engaged with  
38 their infants and toddlers. The use of the existing CHILDES database resulted in a composite  
39 sample that was robust in size and included more than 250,000 words. Consistent with other  
40 studies of parental input (Hart & Risley, 1995), there was considerable variation in the quantity  
41 of input across the 16 mothers drawn from the Brent corpus for analysis in the current study.  
42 Nevertheless, the range in the number of total words (averaging 303.83 to 1177.67 per sample)  
43 was far greater than the range in the number of different words (240.50 to 521.00 per sample).  
44 Mothers with the highest number of total words averaged nearly four times more words than  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 mothers who used the fewest total words, but mothers who produced the greatest number of  
4 different words used just over twice as many different words as those who used the fewest  
5 different words. Less variability in word diversity (i.e., number of different words) relative to  
6 different words. Less variability in word diversity (i.e., number of different words) relative to  
7 word density (i.e., total number of words) has also been reported for parents interacting with  
8 older infants, toddlers, and preschoolers (Rowe, 2012).  
9

10  
11  
12  
13  
14  
15 A substantial proportion of the total words in the composite sample was represented by a  
16 relatively small set of words. The 267 most frequently occurring words comprised approximately  
17 80% of the words used by mothers in their child-directed speech. These findings are consistent  
18 with previous studies that report a core vocabulary of approximately 200 to 300 words  
19 accounting for 80% of the language samples, whether investigating preschoolers (Beukelman et  
20 al., 1989; Marvin et al., 1994), school-age children (Boenisch & Soto, 2015), or adults (Balandin  
21 & Iacono, 1999). The most frequently occurring words were also common to most if not all the  
22 mothers in this sample. Of the 267 high-usage words, 173 were common to all mothers and 248  
23 were common to 13 or more mothers. Higher consistency of use among more frequently  
24 occurring words has also been reported in word studies among preschoolers (Beukelman, et al.,  
25 1989) and adults (Beukelman, Yorkston, Poblete, & Naranjo, 1984).  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40  
41 There was a great deal of overlap between the words mothers in the Brent corpus used  
42 most frequently and those used most frequently by young children in other studies. The greatest  
43 overlap with the mothers occurred with the 23 most frequent words used by toddlers between 24-  
44 36 months of age (Banajee, et al., 2003). The considerable overlap between maternal input just  
45 prior to the one-word stage and children's vocabulary usage at the one- to three-word stage of  
46 development is not surprising. It is to be expected that those words that children encounter more  
47 frequently in interactions with primary caregivers would be more heavily represented among the  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 words used most frequently by young children. When comparing the most frequently used words  
4  
5 by preschoolers (Beukelman et al., 1989; Marvin et al., 1994) with those of mothers from the  
6  
7 Brent corpus, more than half to nearly two-thirds of the most frequently occurring words were  
8  
9 common across studies. Words on the maternal core vocabulary list that did not appear in the  
10  
11 children's most frequent vocabulary lists primarily included the following categories: colloquial  
12  
13 contractions ("wanna", "gotta"), social pragmatic conventions ("thank you", "sorry"),  
14  
15 backchannels ("mhm", "uh-huh"), onomatopoeia ("boom", "whee"), content specific nouns  
16  
17 ("mouth", "outside"), content specific verbs ("read", "walk"), and adjectives ("pretty", "silly").  
18  
19 Thus, while considerable overlap occurs between the most frequently used words in maternal  
20  
21 input and preschooler output, there are still numerous differences between the lists.  
22  
23  
24  
25

### 26 **Implications for Practice**

27  
28 Among children with complex communication needs, vocabulary selection for aided  
29  
30 AAC has almost exclusively been driven by consideration of expressive language needs.  
31  
32 However, receptive language is critical to expressive language development and should play a  
33  
34 role in determining an initial lexicon for parents and clinicians to use when providing aided input  
35  
36 to young children. The results of this study suggest there is a set of frequently occurring words  
37  
38 that mothers use with their prelinguistic children who are typically developing that overlaps with  
39  
40 the words young children who are typically developing use most frequently in their expressive  
41  
42 language. While caution is always advised when applying developmental information from  
43  
44 children with typical development to children with complex communication needs, this study  
45  
46 provides a beginning place for guiding vocabulary selection that is theoretically congruent with a  
47  
48 transactional model of language development.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 This study contributes to the growing literature base of vocabulary studies among various  
4 populations. Several investigations have centered on the words young children with typical  
5 development use most frequently. These productive vocabulary lists are an important source for  
6 guiding vocabulary selection among children who use AAC, with a focus on supporting  
7 expressive language needs (Boenisch & Soto, 2015; Trembath et al., 2007). One significant  
8 finding is the considerable overlap between these lists of the most frequently used words by  
9 children and the words most frequently used by mothers in their child directed speech. If this  
10 overlapping set of words is used in vocabulary selection when designing aided AAC systems for  
11 young children with complex communication needs, it could improve input-output symmetry and  
12 increase opportunities for these children to learn and use graphic symbols to communicate.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

26 It is well established that both the quantity and quality of parent vocabulary input have a  
27 significant influence on the language development of children who use natural speech (Hart &  
28 Risley, 1995; Rowe, 2012). Nevertheless, different aspects of input differentially influence  
29 language acquisition over the course of development, with quantity serving a more influential  
30 role during the emergence of symbolic communication, and diversity serving a more critical role  
31 over the third year of life (Rowe, 2012). This study revealed that, while the mothers did use some  
32 concrete or content specific words (e.g., “kitty”, “book”, “toy”) they were dramatically out-  
33 numbered by words that are used flexibly across a range of contexts and communication partners  
34 (e.g., “you”, “here”, “go”). If vocabulary selection for aided input using graphic symbols with  
35 beginning communicators includes the words adults use most frequently in child-directed speech,  
36 caregivers might provide more meaningful, relevant, and frequent models of graphic symbols use  
37 throughout the child’s day. Therefore, these flexible and non-specific core words may optimally  
38 facilitate high input levels that best support early language development.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 The goal of aided input with graphic symbols, and parental language models in general, is  
4 to provide input that is maximally useful for scaffolding children's behaviors into successful  
5 communication acts, and to build a language system. This means that the most frequent words  
6 used by both mothers and young children may be particularly valuable for supporting both  
7 receptive and expressive language among children with complex communication needs.  
8  
9 Nevertheless, a considerable number of the most frequent words that mothers use with their  
10 prelinguistic infants identified in the present study are not found among the published core  
11 vocabulary lists of preschool children; however, they may have a critical influence on the long-  
12 term communication and language development among children with complex communication  
13 needs (Rowe, 2012). Therefore, including both the overlapping maternal-child core vocabulary  
14 as well as vocabulary that is specific to mothers on the aided AAC systems of beginning  
15 communicators may better support the comprehensive, long-term goals of communication and  
16 language development among children with complex communication needs.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

32  
33 **Iconicity.** One challenge inherent in representing core vocabulary on an aided AAC  
34 system for a beginning communicator is iconicity. Graphic symbols representing content words,  
35 especially nouns, tend to have high levels of iconicity because they can easily represent their  
36 referents, but graphic symbols have much lower levels of iconicity when used to represent more  
37 conceptually-referenced words (Snodgrass et al., 2013). As a result of this lower iconicity,  
38 children may not immediately recognize the meaning of symbols that represent core words  
39 (Light, 2016). Fortunately, young children and beginning communicators who have not yet  
40 developed symbolic communication are less impacted by iconicity than their peers who have an  
41 understanding of symbolic representation (Namy, 2001; Namy et al., 2004). Furthermore,  
42 general experience with pictures, understanding of the words that symbols represent,  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 understanding of the intent of the person using the symbol, and specific instruction in how to use  
4  
5 graphic symbols, have a greater impact than iconicity itself in supporting children in learning to  
6  
7 use graphic symbols (Stephenson, 2009). Given that words classified as core words are used  
8  
9 more frequently than other words, experience can mitigate challenges imposed by limited  
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

### **Limitations and Future Directions**

While this study provides an important first step into better understanding the vocabulary used in child-directed speech, several limitations should be considered when interpreting and applying the results. First, the literature regarding the most frequently used words of young children with typical development is both sparse and dated, with the available lists used in the current comparison ranging from 15- to 29-years-old. Similarly, the maternal input dataset used in the current analysis is dated; however, the focus on word frequency across both child and maternal lists all but eliminates the content words that are subject to variation with time and place. Therefore, it is believed that the current findings provide a meaningful first look at the words mothers use most frequently in their child-directed speech and how those words overlap with the words young children with typical development use in their expressive communication despite the age of the data sets.

A second limitation rests in the use of data sets that involve children with no known disabilities. Differences may exist in the interactions that occur between parents of children with and without disabilities that could influence vocabulary use (Majorano & Lavelli, 2014). In fact, it is well established that maternal input is influenced by child factors such as the child's language ability, intelligibility, cognition, and behaviors (Ciciolla, Gerstein, & Crnic, 2014; Sterling & Warren, 2014; Sterling, Warren, Brady & Felming, 2013). However, finding a way to

1  
2  
3 overcome these differences and minimize the receptive-expressive asymmetry for young children  
4  
5 with complex communication needs was a goal of the current study. As such, it was important to  
6  
7 investigate the words mothers used most frequently in the speech they directed toward their  
8  
9 prelinguistic children without known disabilities. Future research might meaningfully compare  
10  
11 the maternal use of child-directed speech with children with and without communication  
12  
13 disorders, but perhaps it would be more meaningful to compare child-directed speech with and  
14  
15 without access to core vocabulary to support aided input. Furthermore, there may be differences  
16  
17 in vocabulary input between mothers and fathers, as well as among parents of different cultural,  
18  
19 ethnic, and linguistic backgrounds. Future studies should include parents of children with and  
20  
21 without disabilities, who represent diverse ethnic, cultural, economic, educational, and linguistic  
22  
23 backgrounds.  
24  
25  
26  
27

28  
29 This study is also limited by the use of existing data from mother-child dyads and  
30  
31 unrelated preschool-aged children. The children who were the source of the preschool-data in the  
32  
33 current study never interacted with the mothers whose transcripts were analyzed. Future research  
34  
35 should follow parent-child dyads longitudinally to determine how the words parents use in  
36  
37 infancy relate to the words their children use through the preschool years.  
38  
39

### 40 **Conclusion**

41  
42 This study examined the vocabulary used by English-speaking mothers during  
43  
44 interactions with their infants and toddlers who have no known disabilities. Similar to word  
45  
46 frequency studies with other populations, a relatively small set of 267 words accounted for  
47  
48 approximately 80% of the composite sample. The words, common to most if not all of the  
49  
50 mothers in the sample, primarily included words that are flexible across a range of contexts and  
51  
52 partners. Furthermore, more than half of the words mothers used most frequently overlapped  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 with the words unrelated toddlers and preschoolers used most frequently. Therefore, this word  
4  
5 list serves as an important resource for guiding vocabulary selection for caregivers to use during  
6  
7 aided input, and for young children with complex communication needs to have access to use  
8  
9  
10 during early efforts to use graphic symbols to communicate.  
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

For Peer Review Only

### References

- Alant, E., Bornman, J., & Lloyd, L. L. (2006). Issues in AAC research: How much do we really understand. *Disability and Rehabilitation, 28*, 143-150. doi:10.1080/09638280500078000
- Balandin, S., & Iacono, T. (1999). Crews, wusses, and whoppas: Core and fringe vocabularies of Australian meal-break conversations in the workplace. *Augmentative and Alternative Communication, 15*, 95–109. doi:10.1080/07434619912331278605
- Ballin, L., Balandin, S., Stancliffe, R. J., & Togher, L. (2011). Speech-language pathologists' views on mentoring by people who use speech generating devices. *International Journal of Speech-Language Pathology, 13*, 446–457. doi:10.3109/17549507.2011.522254
- Banajee, M., Dicarlo, C., & Stricklin, S. B. (2003). Core vocabulary determination for toddlers. *Augmentative and Alternative Communication, 19*, 67–73. doi:10.1080/0743461031000112034
- Beukelman, D., Jones, R., & Rowan, M. (1989). Frequency of word usage by nondisabled peers in integrated preschool classrooms. *Augmentative and Alternative Communication, 5*, 243–248. doi:10.1080/07434618912331275296
- Beukelman, D. R., & Mirenda, P. (2015). *Augmentative and alternative communication: Supporting children and adults with complex communication needs* (3rd ed.). Baltimore, MD: Paul H. Brookes.
- Beukelman, D. R., Yorkston, K. M., Poblete, M., & Naranjo, C. (1984). Frequency of word occurrence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders, 49*, 360–367. doi:10.1044/jshd.4904.360
- Boenisch, J., & Soto, G. (2015). The oral core vocabulary of typically developing English-speaking school-aged children: Implications for AAC practice. *Augmentative and*

- 1  
2  
3 *Alternative Communication*, 31, 77–84. doi:10.3109/07434618.2014.1001521
- 4  
5 Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary  
6  
7 development. *Cognition*, 81, B33–B44. doi:10.1016/S0010-0277(01)00122-6
- 8  
9  
10 Ciciolla, L., Gerstein, E., & Crnic, K. (2014). Reciprocity among maternal distress, child  
11  
12 behavior, and parenting: Transactional processes and early childhood risk. *Journal of*  
13  
14 *Clinical Child and Adolescent Psychology*, 43, 751-764.  
15  
16 doi:10.1080/15374416.2013.812038
- 17  
18  
19 Deckers, S. R. J. M., Van Zaalen, Y., Van Balkom, H., & Verhoeven, L. (2017). Core  
20  
21 vocabulary of young children with Down syndrome. *Augmentative and Alternative*  
22  
23 *Communication*, 33, 77–86. doi:10.1080/07434618.2017.1293730
- 24  
25  
26 Dennis, A., Erickson, K., & Hatch, P. (2013). The Dynamic Learning Maps core vocabulary:  
27  
28 Overview [technical review]. Retrieved from  
29  
30 <http://www.med.unc.edu/ahs/clds/files/vocabulary-overview>
- 31  
32  
33 Drager, K. R., Light, J. C., Speltz, J. C., Fallon, K. A., & Jeffries, L. Z. (2003). The performance  
34  
35 of typically developing 2.5 year-olds on dynamic display AAC technologies with different  
36  
37 system layouts and language organizations. *Journal of Speech, Language, and Hearing*  
38  
39 *Research*, 46, 298-312. doi: 10.1044/1092-4388(2003/024)
- 40  
41  
42 Geist, L., Hatch, P., & Erickson, K. (2014). Promoting academic achievement for early  
43  
44 communicators of all ages. *SIG 12 Perspectives on Augmentative and Alternative*  
45  
46 *Communication*, 23, 173–181. doi:10.1044/aac23.4.173
- 47  
48  
49 Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experiences of young*  
50  
51 *American children*. Baltimore, MD: Paul H. Brookes.
- 52  
53  
54 Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early  
55  
56  
57  
58  
59  
60

vocabulary development via maternal speech. *Child Development*, 74, 1368–1378.

doi:10.1111/1467-8624.00612

Hurtado, N., Marchman, V.A., & Fernald, A. (2008). Does input influence uptake? Links

between maternal talk, processing speed and vocabulary size in Spanish-learning children.

*Developmental Science*, 11, F31–F39. doi:10.1111/j.1467-7687.2008.00768.x

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary

growth: Relation to language input and gender. *Developmental Psychology*, 27, 236–248.

doi:10.1037/0012-1649.27.2.236

Light, J., Drager, K., McCarthy, J., Mellott, S., Millar, D., Parrish, C., ... Welliver, M. (2004).

Performance of typically developing four- and five-year-old children with AAC systems

using different language organization techniques. *Augmentative and Alternative*

*Communication*, 20, 63–68. doi:10.1080/07434610410001655553

Lloyd, L. L., & Fuller, D. R. (1990). The role of iconicity in augmentative and alternative

communication-symbol learning. In W. I. Fraser (Ed.), *Key issues in mental retardation*

*research* (pp. 295–306). London, UK: Routledge.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale, N.J.:

Lawrence Erlbaum.

Majorano, M., & Lavelli, M. (2014). Maternal input to children with specific language

impairment during shared book reading: Is mothers' language in tune with their children's

production? *International Journal of Language and Communication Disorders*, 49, 204-

214. doi:10.1111/1460-6984.12062

Marvin, C., Beukelman, D., & Bilyeu, D. (1994). Vocabulary-use patterns in preschool children:

Effects of context and time sampling. *Augmentative and Alternative Communication*, 10,

224–236. doi:10.1080/07434619412331276930

McLean, J., & McLean, L. (1999). *How children learn language: A guide for professionals in early childhood or special education*. San Diego, CA: Singular Publishing Group.

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56).

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-old's mapping of nonverbal symbols to object categories. *Infancy*, 2, 73–86.

doi:10.1207/S15327078IN0201\_5

Namy, L. L., Campbell, A. L., & Tomasello, M. (2004). The changing role of iconicity on non-verbal symbol learning: A U-shaped trajectory in the acquisition of arbitrary gestures.

*Journal of Cognition and Development*, 5, 37–57. doi:10.1207/s15327647jcd0501\_3

O'Neil, T., Light, J., & Pope, A. (2018). Effects of interventions that include aided augmentative and alternative communication input on the communication of individuals with complex communication needs: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 61, 1743-1765. doi:10.1044/2018\_JSLHR-L-17-0132

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development* 76, 763-782.

doi:10.1111/1467-8624.00498-i1

Poll, G. H. (2011). Increasing the odds: Applying emergentist theory in language intervention.

*Language, Speech, and Hearing Services in Schools*, 42, 580–591. doi:10.1044/0161-

1461(2011/10-0041)

Python Software Foundation. (2016). *Python language reference* (Version 3.6). Wilmington, DE.

Available at <http://www.python.org>



- 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
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*, 1762–1774.  
doi:10.1111/j.1467-8624.2012.01805.x.A
- Schlosser, R. W. (2003). Selecting graphic symbols for an initial requesting lexicon. In R. W. Schlosser (Ed.), *The efficacy of augmentative and alternative communication: Towards evidence-based practice* (pp. 347–401). New York: Academic Press.
- Sennott, S. C., Light, J. C., & McNaughton, D. (2016). AAC modeling intervention research review. *Research and Practice for Persons with Severe Disabilities, 41*, 101–115.  
doi:10.1177/1540796916638822
- Simcock, G., & DeLoache, J. (2006). Get the picture? The effects of iconicity on toddlers' reenactment from picture books. *Developmental Psychology, 42*, 1352–1357. doi: 10.1037/0012-1649.42.6.1352
- Snodgrass, M. R., Stoner, J. B., & Angell, M. E. (2013). Teaching conceptually referenced core vocabulary for initial augmentative and alternative communication. *Augmentative and Alternative Communication, 29*, 322–33. doi:10.3109/07434618.2013.848932
- Stephenson, J. (2009). Iconicity in the development of picture skills: Typical development and implications for individuals with severe intellectual disabilities. *Augmentative and Alternative Communication, 23*, 187-201. doi:10.1080/07434610903031133
- Sterling, A. M., & Warren, S. F. (2014). Maternal responsivity in mothers of young children with Down syndrome. *Developmental Neurorehabilitation, 17*, 306-317.  
doi:10.3109/17518423.2013.772671
- Sterling, A. M., Warren, S. F., Brady, N., & Fleming, K. (2013). Influences on maternal responsivity in mothers of children with fragile X syndrome. *American Journal on*

1  
2  
3 *Intellectual and Developmental Disabilities*, 118, 310-326. doi:10.1352/1944-7558-  
4  
5 188.4.310

6  
7  
8 Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*.  
9  
10 Cambridge, MA: Harvard University Press.

11  
12 Trembath, D., Balandin, S., & Togher, L. (2007). Vocabulary selection for Australian children  
13  
14 who use augmentative and alternative communication. *Journal of Intellectual and*  
15  
16 *Developmental Disability*, 32, 291–301. doi:10.1080/13668250701689298

17  
18  
19 Trudeau, N., Cleave, P., & Woelk, E. (2003). Using augmentative and alternative  
20  
21 communication approaches to promote participation of preschoolers during book reading: a  
22  
23 pilot study. *Child Language Teaching and Therapy*, 19, 181–210.  
24  
25 doi:10.1191/0265659003ct25oa

26  
27  
28 Van Tatenhove, G. M. (2009). Building language competence with students using AAC devices:  
29  
30 Six challenges. *SIG 12 Perspectives on Augmentative and Alternative Communication*, 18,  
31  
32 38–47. doi:10.1044/aac18.2.38

33  
34  
35 von Tetzchner, S. (2015). The semiotics of aided language development. *Cognitive*  
36  
37 *Development*, 36, 180–190. doi:10.1016/j.cogdev.2015.09.009

38  
39  
40 Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency*  
41  
42 *guide*. Brewster, NY: Touchstone Applied Science.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Running Header: MOST FREQUENT WORDS CHILD-DIRECTED SPEECH

Table 1

*Language Analysis of Maternal Transcripts*

Mother	Mean number of utterances	Mean length of utterance: Words	Mean length of utterance: Morphemes	Mean number of different words	Mean number of total words	Mean type-token ratios
1	933.17	3.75	4.41	441.67	3791.33	0.12
2	922.83	3.78	4.29	477.50	3685.83	0.13
3	762.00	3.70	4.16	400.00	3011.67	0.13
4	1032.50	2.89	3.28	335.83	3228.33	0.11
5	1062.50	3.33	3.92	417.17	3878.33	0.11
6	524.50	2.81	3.18	273.00	1573.83	0.18
7	303.83	3.89	4.48	240.50	1273.00	0.19
8	436.00	3.06	3.47	254.33	1431.17	0.18
9	967.33	3.34	3.86	383.17	3490.33	0.11
10	1177.67	3.20	3.72	456.50	4071.50	0.11
11	595.17	2.95	3.29	278.83	1859.33	0.17
12	753.67	3.78	4.27	420.17	3019.50	0.14
13	860.83	3.44	3.95	359.83	3192.67	0.11
14	793.00	3.58	4.11	378.67	3129.00	0.13
15	914.83	3.63	4.21	521.00	3570.17	0.15
16	744.83	3.43	3.86	443.00	2767.17	0.16
	799.04	3.41	3.90	380.07	2935.82	0.14

## Running head: MOST FREQUENT WORDS CHILD-DIRECTED SPEECH

Table 2

*Percent of Total Composite and Mean Commonality Score for Varying Sized Vocabulary Lists of Mothers' Most Frequently Used Words*

Set of most frequently occurring words	Percentage of total composite sample of words	Mean commonality score
1 - 25	34.49	16
1 - 50	49.93	15.9
1 - 75	59.25	15.93
1 - 100	65.02	15.82
1 - 125	68.74	15.74
1 - 150	71.73	15.67
1 - 175	74.22	15.53
1 - 200	76.18	15.46
1 - 225	77.88	15.39
1 - 250	79.28	15.21
1 - 267	80.22	15.11

## Running head: MOST FREQUENT WORDS CHILD-DIRECTED SPEECH

Table 3

*Percent Overlap between Core Vocabulary Studies of Young Children and Varying Sized Vocabulary Lists*

Set of mothers' most frequently occurring words	Percentage overlap with Banajee et al. (2003) core words ( $n = 23$ )	Percentage overlap with Beukelman et al. (1989) core words ( $n = 238$ )	Percentage overlap with Marvin et al. (1994) core words ( $n = 324$ )
1 - 25	56.52	10.08	7.10
1 - 50	60.87	15.96	13.58
1 - 75	78.26	24.79	20.37
1 - 100	86.98	32.35	26.23
1 - 125	86.96	38.24	32.72
1 - 150	86.96	44.12	38.27
1 - 175	86.96	48.32	41.98
1 - 200	86.96	52.94	46.60
1 - 225	86.96	56.72	51.54
1 - 250	91.30	60.92	54.01
1 - 267	91.30	64.71	58.02

Running head: MOST FREQUENT WORDS CHILD-DIRECTED SPEECH

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

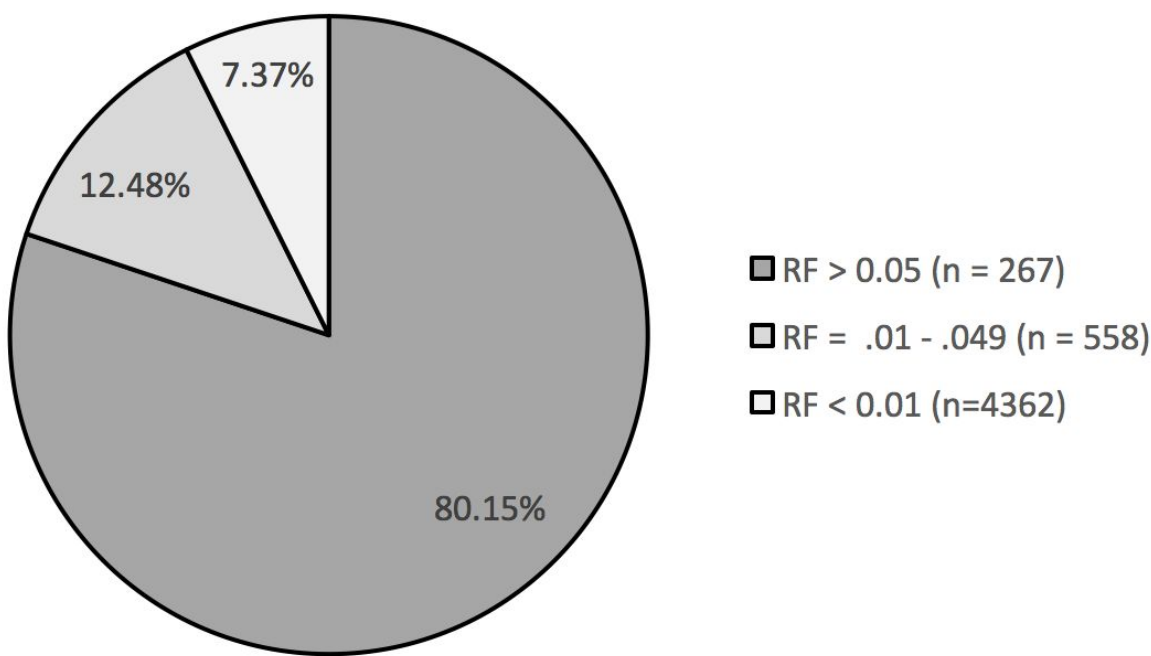


Figure 1. Cumulative frequency for words with different ranges of relative frequency (RF)

Running head: CORE VOCABULARY MATERNAL INPUT

1

## Appendix A

Frequency, Relative Frequency, and Commonality of Maternal Core Vocabulary Words and their Overlap with Core Vocabulary Studies of Young Children

	Word	Frequency count	Composite relative frequency (%)	Commonality score	Banajee et al. (2003) core words ( <i>n</i> = 23)	Beukelman et al. (1989) core words ( <i>n</i> = 238)	Marvin et al. (1994) core words ( <i>n</i> = 324)
1	you	15219	5.9108	16	1	1	1
2	the	6848	2.6596	16	1	1	1
3	a	4655	1.8079	16	1	1	0
4	that	4651	1.8064	16	1	1	1
5	it	4624	1.7959	16	1	1	1
6	your	3759	1.4599	16	0	1	1
7	I	3412	1.3252	16	1	1	1
8	here	3123	1.2129	16	1	1	1
9	to	2971	1.1539	16	0	1	1
10	go	2956	1.1481	16	1	1	1
11	no	2921	1.1345	16	1	1	1
12	on	2918	1.1333	16	1	1	1
13	and	2887	1.1213	16	0	1	1
14	what	2795	1.0855	16	1	1	1
15	is	2646	1.0277	16	1	1	1
16	in	2397	0.9309	16	1	1	1
17	oh	2339	0.9084	16	0	1	1
18	this	2303	0.8944	16	0	1	1
19	there	2250	0.8739	16	0	1	1
20	okay	2213	0.8595	16	0	1	1
21	get	2210	0.8583	16	0	1	1
22	see	2183	0.8478	16	0	1	1
23	come	2148	0.8342	16	0	1	1
24	yeah	2122	0.8241	16	0	0	0

## Running head: CORE VOCABULARY MATERNAL INPUT

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

25	do	2075	0.8059	16	0	1	1
26	read	2074	0.8055	14	0	0	0
27	up	2063	0.8012	16	0	1	1
28	voice	2034	0.7900	15	0	0	0
29	don't	1978	0.7682	16	0	0	1
30	are	1946	0.7558	16	0	1	1
31	gonna	1794	0.6968	16	0	0	0
32	we	1793	0.6964	16	0	1	1
33	can	1752	0.6804	16	0	1	1
34	say	1744	0.6773	16	0	0	1
35	that's	1724	0.6696	16	0	0	1
36	sung	1641	0.6373	16	0	0	0
37	look	1552	0.6028	16	0	1	1
38	put	1527	0.5931	16	0	1	1
39	me	1491	0.5791	16	0	1	1
40	good	1478	0.5740	16	0	1	1
41	mommy	1442	0.5600	14	0	0	0
42	it's	1424	0.5531	16	0	0	1
43	one	1415	0.5496	16	0	1	1
44	want	1331	0.5169	16	1	1	1
45	you're	1311	0.5092	16	0	0	1
46	with	1309	0.5084	16	0	1	1
47	let's	1305	0.5068	16	0	0	1
48	wanna	1283	0.4983	16	0	0	0
49	for	1168	0.4536	16	0	1	1
50	like	1167	0.4532	16	0	1	1
51	huh	1153	0.4478	16	0	0	0
52	little	1124	0.4365	16	0	1	1
53	not	1109	0.4307	16	0	1	1
54	baby	1088	0.4226	16	0	1	1
55	got	1087	0.4222	16	0	1	0
56	all	1080	0.4195	16	0	1	1
57	some	1047	0.4066	16	1	1	1



Running head: CORE VOCABULARY MATERNAL INPUT

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

58	down	1044	0.4055	16	0	1	1
59	have	1044	0.4055	16	0	1	1
60	of	1028	0.3993	16	0	1	1
61	now	1006	0.3907	16	0	1	0
62	know	950	0.3690	16	0	1	1
63	my	944	0.3666	16	1	1	1
64	where	931	0.3616	16	0	1	1
65	at	925	0.3593	16	0	1	1
66	out	909	0.3530	16	1	1	1
67	whispered	894	0.3472	16	0	0	0
68	did	869	0.3375	16	0	1	1
69	I'm	868	0.3371	16	0	0	1
70	so	860	0.3340	16	0	1	1
71	just	853	0.3313	16	0	1	1
72	back	843	0.3274	16	0	1	1
73	yes	798	0.3099	16	1	1	1
74	what's	797	0.3095	16	0	0	1
75	play	759	0.2948	16	0	1	1
76	uh-oh	739	0.2870	16	0	0	0
77	he	732	0.2843	16	0	1	1
78	boy	703	0.2730	14	0	0	1
79	going	701	0.2723	16	0	1	1
80	hi	701	0.2723	16	0	1	1
81	be	689	0.2676	16	0	1	1
82	them	656	0.2548	16	0	1	1
83	can't	646	0.2509	16	0	0	0
84	doing	627	0.2435	16	0	1	1
85	more	618	0.2400	16	1	1	1
86	right	612	0.2377	16	0	1	1
87	where's	589	0.2288	14	0	0	0
88	eat	580	0.2253	16	0	1	1
89	hey	549	0.2132	16	0	1	0
90	over	548	0.2128	16	0	1	1

Running head: CORE VOCABULARY MATERNAL INPUT

1								
2								
3	91	think	540	0.2097	16	0	1	0
4	92	off	531	0.2062	16	1	1	1
5	93	how	530	0.2058	16	0	1	1
6	94	big	527	0.2047	15	0	1	1
7	95	take	516	0.2004	16	0	1	0
8	96	there's	514	0.1996	16	0	0	1
9	97	kitty	510	0.1981	10	0	0	0
10	98	mommy's	501	0.1946	15	0	0	0
11	99	ball	501	0.1946	15	0	1	1
12	100	was	488	0.1895	16	0	1	1
13								
14	101	let	472	0.1833	16	0	1	1
15	102	book	468	0.1818	15	0	0	0
16	103	gotta	468	0.1818	16	0	0	0
17	104	girl	448	0.1740	9	0	1	1
18	105	too	445	0.1728	16	0	1	1
19	106	mama	441	0.1713	14	0	0	0
20	107	well	415	0.1612	16	0	0	1
21	108	two	409	0.1588	16	0	1	1
22	109	thank you	404	0.1569	16	0	0	0
23	110	about	397	0.1542	16	0	0	1
24	111	he's	393	0.1526	16	0	0	1
25	112	again	392	0.1522	16	0	1	1
26	113	him	392	0.1522	16	0	1	1
27	114	why	388	0.1507	15	0	1	1
28	115	but	386	0.1499	16	0	1	1
29	116	they	384	0.1491	16	0	1	1
30	117	stop	371	0.1441	14	0	1	1
31	118	those	369	0.1433	15	0	1	1
32	119	hold	366	0.1421	16	0	0	1
33	120	give	362	0.1406	16	0	1	1
34	121	if	354	0.1375	16	0	0	1
35	122	ready	353	0.1371	16	0	1	1
36	123	very	349	0.1355	16	0	0	1
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47								

## Running head: CORE VOCABULARY MATERNAL INPUT

5

1							
2							
3	124	water	345	0.1340	16	0	1
4	125	here's	341	0.1324	16	0	1
5							
6	126	alright	336	0.1305	16	0	0
7	127	his	329	0.1278	16	0	1
8		somethin					
9	128	g	329	0.1278	16	0	1
10	129	need	327	0.1270	16	0	1
11	130	hello	326	0.1266	14	0	1
12	131	bye-bye	323	0.1254	15	0	0
13	132	mhm	322	0.1251	14	0	0
14	133	done	320	0.1243	16	0	1
15	134	she	315	0.1223	15	0	1
16	135	I'll	315	0.1223	16	0	1
17	136	sit	314	0.1220	16	0	1
18	137	have_to	313	0.1216	16	0	0
19	138	these	308	0.1196	15	0	1
20	139	we'll	305	0.1185	16	0	1
21	140	we're	305	0.1185	16	0	1
22	141	sweetie	298	0.1157	13	0	0
23	142	because	296	0.1150	16	0	1
24	143	nose	289	0.1122	16	0	0
25	144	turn	287	0.1115	16	0	1
26	145	getting	284	0.1103	16	0	1
27	146	her	280	0.1087	15	0	1
28	147	three	279	0.1084	14	0	1
29	148	open	277	0.1076	14	0	1
30	149	away	274	0.1064	16	0	1
31	150	tickle	274	0.1064	13	0	0
32							
33	151	mouth	270	0.1049	16	0	0
34	152	who	270	0.1049	16	0	1
35	153	peekaboo	267	0.1037	13	0	0
36	154	way	266	0.1033	16	0	1
37	155	whoa	265	0.1029	13	0	0
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							

## Running head: CORE VOCABULARY MATERNAL INPUT

1								
2								
3	156	time	263	0.1021	16	0	1	1
4	157	nice	260	0.1010	16	0	0	1
5	158	throw	259	0.1006	16	0	0	0
6	159	trying	254	0.0986	16	0	1	1
7	160	make	253	0.0983	16	0	1	1
8	161	bite	249	0.0967	14	0	0	1
9	162	lemme	245	0.0952	16	0	0	0
10	163	walk	240	0.0932	16	0	0	0
11	164	stuff	239	0.0928	16	0	0	1
12	165	head	239	0.0928	16	0	0	1
13	166	then	236	0.0917	16	0	1	1
14	167	oops	234	0.0909	12	0	0	0
15	168	yummy	234	0.0909	11	0	0	0
16	169	does	233	0.0905	15	0	1	1
17	170	will	232	0.0901	16	0	1	1
18	171	bear	230	0.0893	13	0	0	0
19	172	daddy	230	0.0893	12	0	0	0
20	173	boom	228	0.0886	10	0	0	0
21	174	feet	227	0.0882	15	0	1	0
22	175	wait	222	0.0862	15	0	0	1
23	176	oh	221	0.0858	16	0	0	0
24	177	love	221	0.0858	13	0	0	0
25	178	hands	220	0.0854	16	0	1	1
26	179	try	219	0.0851	16	0	1	1
27	180	clean	219	0.0851	16	0	1	0
28	181	ow	218	0.0847	14	0	0	0
29	182	when	210	0.0816	16	0	1	1
30	183	thing	207	0.0804	16	0	0	1
31	184	whee	205	0.0796	10	0	0	0
32	185	doggie	205	0.0796	12	0	0	0
33	186	still	205	0.0796	15	0	0	1
34	187	didn't	204	0.0792	16	0	0	1
35	188	outside	203	0.0788	14	0	0	0
36								
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47								

## Running head: CORE VOCABULARY MATERNAL INPUT

7

1							
2							
3	189	hair	203	0.0788	16	0	1
4	190	whatcha	199	0.0773	16	0	0
5	191	other	197	0.0765	16	0	1
6	192	around	192	0.0746	15	0	1
7	193	hand	192	0.0746	16	0	1
8	194	diaper	190	0.0738	13	0	0
9	195	juice	190	0.0738	13	0	1
10	196	foot	189	0.0734	16	0	1
11	197	another	187	0.0726	16	0	1
12	198	pretty	187	0.0726	16	0	0
13	199	from	187	0.0726	16	0	1
14	200	or	186	0.0722	16	0	1
15	201	hear	186	0.0722	16	0	1
16	202	has	186	0.0722	16	0	1
17	203	today	186	0.0722	16	0	1
18	204	man	184	0.0715	15	0	1
19	205	stand	184	0.0715	16	0	0
20	206	minute	183	0.0711	15	0	0
21	207	yum	183	0.0711	10	0	0
22	208	would	181	0.0703	16	0	1
23	209	goes	180	0.0699	14	0	1
24	210	find	179	0.0695	16	0	1
25	211	toy	176	0.0684	15	0	0
26	212	looking	175	0.0680	16	0	1
27	213	watch	172	0.0668	16	0	1
28	214	close	172	0.0668	14	0	0
29	215	really	171	0.0664	16	0	1
30	216	kiss	171	0.0664	16	0	0
31	217	had	171	0.0664	16	0	1
32	218	ya	171	0.0664	13	0	0
33	219	yellow	169	0.0656	13	0	0
34	220	drink	167	0.0649	15	0	1
35	221	push	166	0.0645	14	0	1
36							
37							
38							
39							
40							
41							
42							
43							
44							
45							
46							
47							

## Running head: CORE VOCABULARY MATERNAL INPUT

1								
2								
3	222	clap	165	0.0641	12	0	0	0
4	223	face	165	0.0641	16	0	0	1
5	224	better	164	0.0637	16	0	0	1
6	225	wash	164	0.0637	14	0	1	0
7								
8	226	else	163	0.0633	15	0	1	1
9	227	door	163	0.0633	14	0	1	1
10	228	isn't	156	0.0606	13	0	0	1
11	229	honey	156	0.0606	11	0	0	0
12	230	four	155	0.0602	14	0	1	0
13	231	tell	155	0.0602	16	0	1	1
14	232	bottle	154	0.0598	12	0	0	0
15	233	pull	154	0.0598	15	0	0	0
16	234	who's	152	0.0590	13	0	0	0
17	235	coming	151	0.0586	16	0	1	0
18	236	silly	150	0.0583	11	0	0	0
19	237	funny	149	0.0579	16	0	0	0
20	238	chair	148	0.0575	15	0	1	1
21	239	fish	147	0.0571	10	0	1	0
22	240	cold	147	0.0571	14	0	0	0
23	241	teeth	147	0.0571	9	0	0	0
24	242	help	146	0.0567	14	1	1	1
25	243	yourself	145	0.0563	16	0	0	0
26	244	fall	145	0.0563	15	0	1	1
27	245	spoon	143	0.0555	14	0	0	0
28	246	bunny	142	0.0551	13	0	0	0
29	247	they're	142	0.0551	13	0	0	1
30	248	pooh	142	0.0551	10	0	0	0
31	249	out_of	142	0.0551	15	0	0	0
32	250	bit	141	0.0548	15	0	1	0
33	251	toys	141	0.0548	15	0	0	1
34	252	yay	140	0.0544	8	0	0	0
35	253	won't	139	0.0540	15	0	0	1
36	254	much	138	0.0536	15	0	0	1
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47								

## Running head: CORE VOCABULARY MATERNAL INPUT

9

255	any	138	0.0536	15	0	1	1
256	food	137	0.0532	15	0	1	0
257	could	137	0.0532	16	0	1	1
258	green	135	0.0524	14	0	1	1
259	car	135	0.0524	13	0	1	1
260	blue	134	0.0520	11	0	1	1
261	things	133	0.0517	14	0	1	1
262	sh	133	0.0517	9	0	0	0
263	went	133	0.0517	15	0	1	1
264	please	133	0.0517	14	0	1	1
265	sorry	130	0.0505	15	0	0	0
266	people	129	0.0501	15	0	0	1
267	doesn't	129	0.0501	13	0	0	1