



Published in final edited form as:

Child Dev. 2018 July ; 89(4): 1247–1267. doi:10.1111/cdev.12731.

Quantitative linguistic predictors of infants' learning of specific English words

Daniel Swingley and Colman Humphrey

University of Pennsylvania

Abstract

To evaluate which features of spoken language aid infant word learning, a corpus of infant-directed speech (Brent & Siskind, 2001) was characterized on several linguistic dimensions and statistically related to the infants' vocabulary outcomes word-by-word. Comprehension (at 12, 15 months) and production (15 months), were predicted by frequency, frequency of occurrence in one-word utterances, concreteness, utterance length, and typical duration. These features have been proposed to influence learning before, but here their relative contributions were measured. Mothers' data predicted learning in their own children better than in other children; thus, vocabulary is measurably aligned within families. These analyses provide a quantitative basis for claims concerning the relevance of several properties of maternal English speech in facilitating early word learning.

Keywords

language acquisition; lexical development; word learning; infant language

Infants begin their language-learning careers hearing a great deal of speech and understanding none of it. As they listen over the course of their first months, words, or something like words, begin to emerge from the phonetic flow, as familiar and recurring bits of language. By six months, some of these familiar speech sequences also become invested with meaning, and so they are not just sounds, but sounds that have a connection to the outside world of daddies and milk-bottles and hands (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012). And then, within the following year, typical toddlers develop a vocabulary of hundreds of words. How does this happen?

This question has been of interest to science for over a century (e.g., Taine, 1877). Learning words is a major part of language development and building the vocabulary is a significant educational outcome, so the normative developmental course has been studied a great deal, especially in schoolchildren (e.g., Whipple, 1925). In recent years, increasing attention has been devoted to the earliest stages of word learning. Understanding word learning is part of the scientific community's attempt to characterize the human capacity for language (e.g., Landau & Gleitman, 1985; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002,

and many others). The question of how children learn words has also taken on new urgency because of the possibility that intervention might elevate the early language skills of children in disadvantaged social groups, and thereby (it is hoped) reduce social inequalities in school readiness and performance (e.g., Cates, Weisleder, & Mendelsohn, 2016; Hart & Risley, 1995; Fernald, Marchman, & Weisleder, 2013).

Early word learning has been studied in two ways: by teaching children words in the lab, to characterize the moment of learning, and by surveying children's knowledge through parent report or direct observation. Generally speaking, experiments measure processes, and observational studies measure outcomes. These approaches have been complementary in their conclusions. Experiments have taught us, for example, that young children use several forms of linguistic and social context to narrow down words' meanings (e.g., Bloom, 2000; Gleitman, Cassidy, Nappa, & Papafragou, 2005; Markman & Wachtel, 1988; Yuan & Fisher, 2009), that children's intuitions about categories determine their extension of words to new instances (e.g., Markman, 1989; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002) and that phonetic aspects of words affect learning (e.g., Johnson & Jusczyk, 2001; Storkel, 2001).

Observational studies, including large-scale inventories of children's vocabulary composition, have revealed broad individual differences in vocabulary size, some of which are linked to demographic factors like socioeconomic status (Fenson et al., 1994; Hoff, 2006) and some of which are associated with quantitative properties of the language environment, such as quantity and diversity of child-directed conversation (e.g., Newman, Rowe, & Bernstein-Ratner, 2015; Rowe, 2012; Weisleder & Fernald, 2013). Across children, there is considerable range in the sorts of words children tend to learn, but also consistent trends, such as an eventual predominance of nominals, and an emphasis on words that bear on children's immediate interests (e.g., Nelson, 1973).

In the psycholinguistic literature, individual word-teaching experiments tend to focus on whether or not children can take advantage of one or two sources of information in learning one word, when that information is provided in a clear, concentrated form. The experiments provide less information about linguistic cues' relative importance in the day-to-day business of natural word learning. Observational studies of vocabulary development, as opposed to experiments, often take in the bigger picture, but rarely attempt to explain what leads to the learning of particular words, focusing instead on measures like total word counts in children's vocabularies.

The goal of the present study was to provide a kind of bridge between these two sorts of research program. The current work is one of a few such attempts to approach the problem from different angles, taking elements of the experimental tradition but using more naturalistic learning situations and larger datasets. For example, in a recent study Cartmill et al. (2013) drew from a giant set of longitudinal videorecordings of parent-child interaction to create a database of word-use vignettes. A sample of adults attempted to identify the referents of parents' words in these video clips, without hearing the words themselves. Correlational analyses showed that the more successful the adult subjects were at identifying a parent's meaning, the larger the vocabulary of these parents' children three years later.

This effect was separate from an additional, significant relation between the sheer amount parents talked, and children's vocabulary size at 4.5 years. This study shows that parents differ in measurable ways in how they offer language experience to their children, and that features of this experience relate consistently to learning outcomes among children.

Other studies have focused more on individual words than individual children, typically inferring rather than measuring "input" features of the words (though see Roy, Frank, DeCamp, Miller, & Roy, 2015). Braginsky, Yurovsky, Marchman, and Frank (2016) examined parent-report data from the MacArthur Communicative Development Inventory (MCDI; Fenson et al., 1994), predicting the age at which children would be expected to know a word based on the likely conditions of occurrence of those words, where those conditions were estimated using other databases. Braginsky et al. found that across several languages, earlier learning of words was predicted by frequency (infants learn more common words), "babiness" (infants learn more words that adults think of as baby-related), concreteness (infants learn words for more concrete concepts), and mean length of utterance (infants learn words that tend to occur in shorter sentences, an effect particularly strong in English).

The present study was similar to the Braginsky et al. (2016) work, with the important difference that the exposure characteristics of each word were measured for the same children whose outcomes were being measured. Thus, when evaluating a predictor such as word frequency, the measurement in question was derived from the speech of the mother whose child's word understanding or word production was measured. A second feature unique to the present study was the use of acoustic measurements of maternal speech, as described below.

We examined the role of a wide-ranging set of "input" characteristics in accounting for children's learning of a large and diverse set of specific words. This was done by following up on the pioneering work of Brent and Siskind (2001). Brent and Siskind (2001; hereafter B&S) proposed that one of the most important determinants of whether young children would learn a word is how often they hear that word in isolation: that is, its frequency of occurrence in one-word sentences. To test this hypothesis, Brent and Siskind created, transcribed, and analyzed recordings of 8 mothers speaking to their infants as the infants grew from 9 to 15 months of age. Parents completed a vocabulary checklist (the MacArthur Communicative Development Inventory; Fenson et al., 1994) at 12, 15, and 18 months. These checklists, and also children's attested productive vocabularies as captured in the recordings made at 14–15 months, provided each child's outcome data. The two predictors in B&S's analysis were (a) the frequency with which a word appeared in each child's corpus sample (based on the transcripts of each child's mother when the child was between 9 and 12 months of age); and (b) the frequency with which each word appeared in isolation.

Brent and Siskind found that total frequency, i.e., the overall number of times that a given mother used a particular word, did not predict whether her child would use that word. However, the number of times she said a given word *in isolation* was a significant predictor of which words a given child knew, whether the outcome measure was comprehension vocabulary (from the CDI at 12 months; B&S footnote 3), production vocabulary on the CDI

at 15 months, or child production within the recordings. The authors suggested that a rate-limiting factor in infant word learning is therefore the *segmentation* of words from their phonetic context, a problem that is presumably solved in the case of words having no phonetic context at all. B&S further proposed that words discovered by being heard in isolation could then go on to assist in the discovery of other words using a kind of “remainder-is-a-new-potential-word” assumption (formalized in Brent & Cartwright, 1996). Certainly the finding that overall frequency, as opposed to isolated-word frequency, was of no predictive power whatever, is quite striking, suggesting that pulling a word out of its context is a formidable challenge and a significant impediment to word discovery at very young ages.

The present study re-examined this conclusion using the B&S dataset, but including a much wider range of predictors. Our initial motivation was the concern that isolated-word frequency might in fact be a proxy for something else, such as conceptual simplicity, syntactic simplicity, utterance-final sentence position, or exaggerated phonetic duration. Because in principle any of these features might be correlated with a tendency to occur in isolation, one cannot safely conclude that it was isolated-word frequency itself that was responsible for B&S’s learning outcomes.

Examining this question more fully leads naturally to a broader examination of which features of words (or children’s experience with words) lead to learning. Some words might be intrinsically easier to learn, because they are short, or phonologically simple, or used to denote concepts that children grasp easily. Some words might tend to be used very often, or used in short sentences, or with phonetic enhancement (“hyperarticulation”). Features like these have been linked to word learning in observational studies (Braginsky et al., 2016; Roy et al., 2015; Stokes, 2015) or to word recognition in experiments (Fernald, McRoberts, & Swingley, 2001; Song, Demuth, & Morgan, 2010). By measuring and then entering such factors into regression models, it was possible to assess their predictive power in the unique Brent dataset, and determine which features of words are most strongly associated with their acquisition by young children. Children’s own phonological productions were also examined for their possible predictive value (Vihman & Croft, 2007). In addition, a supplementary analysis examined the degree of alignment between mother and child in the specific words children knew, showing that a given mother’s speech relates much better to the vocabulary outcome of her own child than to other children.

Methods

Preparation of the corpora

Corpus materials were downloaded from the Childes repository (MacWhinney, 2000) in 2005. Significant editing was undertaken on the text transcripts. For example, sentences transcribed with “xxx” or “www” (conventional tags indicating untranscribed speech) were removed, as were sentences tagged as sung, whispered, or read; annotator comments were removed; brackets indicating conversational overlap were removed. The 169 compounds indicated with the “+” character, such as “kitty+cat” and “rocking+chair” were made either one word (e.g., “sleepyhead”, “lawnmower”) or, less often, split (e.g., “sewing machine”, “swimming pool”) based on the author’s intuitions about the words’ likely prosody and the

separate words' independent presence in the corpus (1013 tokens). Punctuation was removed and weird spellings were corrected. Words that were transcribed in parentheses to signal that they were not actually spoken were removed. Then all sentences not marked as maternal ("MOT") were excised. This yielded a corpus of about 130,000 utterances and about 455,000 words (not including punctuation).

Brent and Siskind's Chiles database included recordings and transcripts of 16 children's language environments (together with their CDIs), but B&S narrowed the sample to eight of the children for analysis. Their report does not list which children these were, but the eight children code-named c1, d1, f1, i1, s1, s2, v1, and v2 in the full set of corpora seemed to match the descriptions provided in B&S's Tables 1 and 2, and were selected for the present study. Recordings had been made in 1996 and 1997, except for child v2 who was recorded in 2000. Together the total corpus contained about 83,900 utterances and about 300,000 words.

A summary of the number of words each mother said is given in Table 1.

Predictors

Features of each mother's language were measured from their recordings and transcripts, focusing on the words that appear on the Infant CDI. The predictors tested were: *word duration* (did the mother tend to say the word with a greater duration than would be expected given its length?); *total frequency* of occurrence; *frequency in isolation*, *frequency utterance-finally*, *frequency utterance-medially*; *MLU* (mean word length of the utterances in which the mother used the word); and *phonotactic probability*, which was computed in 3 different ways as described later.

These measures were computed over lemmas, not types, which meant that instances such as "broke" and "broken", or "(ba)nana" and "bananas," and so forth, were counted together. This step is appropriate because parents filling out the CDI probably respond to the lemma and not just the specific form of each listed word. A table linking forms to CDI lemmas was constructed by hand (with the aid of the corpus' morphological annotation, i.e. the "mor%" tier), covering every token in the corpus. From the lemmas, the statistics were computed as follows.

Word duration ratio—A soundfile clip for each utterance in the corpus was extracted from the session recordings. The original transcripts provide onset and offset times for each utterance; these are mostly correct, and formed the basis for the extractions.

To model the phonetics of the spoken words, canonical pronunciations were retrieved from the CMU pronouncing dictionary (version 0.7a; <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>). When the dictionary gave pronunciation variants, the first was used. About 600 words in the corpus were not in the dictionary. A pronunciation was constructed by hand for all of these out-of-dictionary words.

The orthographic transcriptions, soundfiles, and dictionary were fed into the HTK Speech Recognition Toolkit (version 3.4.0, Young et al., 2006) via the Penn Phonetics Lab Forced Aligner front-end (Yuan & Liberman, 2008). These tools use an acoustic model of English

to line up the speech sounds expected from the transcription (given the dictionary) and portions of the soundfiles. Though the alignments are not perfect, they permit estimation of the durations of words and speech sounds.

Because word duration was intended as a measure of hyperarticulation, durations were normalized by the expected durations of the words' canonical forms. (Raw duration does not work because words with many syllables can be long in duration without being hyperarticulated.) Expected durations were computed by tabulating, for each mother, the median duration of each speech sound: that mom's median /b/, her median unstressed /æ/, and so on. The three levels of vowel stress that are represented in the CMU dictionary were maintained for this computation. When a mother used a given sound fewer than 15 times (16% of cases) her value was replaced by the mean of the other mothers' value for that sound, multiplied by the ratio of her overall durations and the other moms' overall durations (to take average speaking rate into account). The expected duration of the word "blue", then, was the sum of a given mother's median /b/, /l/, and /u/. A given instance of that word might be longer or shorter than this expected duration, and thus the ratio of a given instance to its expected duration could be greater than one (longer) or less than one (shorter). The statistic used in the models was the mean ratio of actual to expected duration for each CDI word for each mother. Many CDI words had ratios above one (median = 1.28; 20th percentile = 1.03, 80th percentile = 1.67), reflecting the fact that CDI words tend to be enunciated with longer durations than other words generally.

Not all mothers' corpora included instances of every word on the infant CDI, resulting in exclusion of some child X word instances from the analysis. To ameliorate the effects of these exclusions, missing values of the acoustic duration variables were substituted by mean ratios across the remaining mothers whenever there were at least 3 mothers each contributing at least 2 instances of a word. This made each dataset about 20% larger and helped with model convergence but did not substantively affect the results.

Frequency—Following Brent and Siskind (2001), frequency was computed only for the "early" corpora, i.e. recordings taken before the children's first birthday (a mean of 5.9 70-minute recording sessions per child). Words were considered to be in isolation only if they were the only word in the utterance as transcribed. Isolated words were not also counted as utterance-final, i.e. "utterance-final" means "the last word in an utterance that has more than one word in it."

MLU—Mean length of utterance for each CDI word, computed over lemmas, for each mother. This measure captures the fact that some words tended to be used in relatively long utterances. To help with the sparsity of this measure, each mother's entire corpus was used, not just the earlier corpora. When a word did not appear in a given mother's corpus, the median MLU value was estimated using multiple imputation (Rubin, 1976). This is an estimation technique that guesses at a value by drawing from a distribution based on the mean and standard deviation of the nonmissing values (here, the other mothers' MLU for that word given its semantic category). Because this is a stochastic process, we ran our final models 50 times; our reported coefficients for each parameter are means over these imputed

datasets. (The results were nearly identical when filling in missing MLU data using the average of MLU for non-missing moms.)

Phonotactic probability—Words might be easier or harder to learn as a function of how commonly their sounds occur in the language, and how commonly those sounds occur together. These frequencies were counted over the Brent corpus. *Bigram probability* was calculated by counting the number of times each sound-pair in each word occurred in the corpus (independently of the corpus words the pair appeared in), and computing the mean of the log of these counts. *Bigram type frequency* is the mean of the log frequencies of the words (lemmas) that each bigram in a word appears in. *Neighbors* is the number of phonological neighbors of the word in question, where a “neighbor” is a word the target word could be converted into by adding, subtracting, or substituting a single phone.

Vocal preferences—Children just beginning to say words sometimes favor words containing sounds they have already learned to produce (e.g., Majorano, Vihman, & DePaolis, 2013). To evaluate this in the Brent dataset, all utterances produced by the children (transcribed as “CHI:”) were extracted and retranscribed with attention to the phonetic segments children actually produced. The utterances were not numerous—considering the early (< 12 months) transcripts, children averaged 26 utterances (range, 1–65). There was evidence for some children using a *vocal motor scheme* as described by Vihman (e.g., McCune & Vihman, 2001): 93% of child d1’s consonants (39/42) were [m], and 62% of child s1’s consonants (15/24) were [l]. But the other six children were transcribed as saying no consonants more than 10 or 15 times, and none in conspicuously high proportions. Thus, we were not able to characterize most children as displaying vocal-motor schemes and did not evaluate this variable as a predictor.

Concreteness—This predictor was included to help account for the fact that children tend to learn words for concrete things earlier or more readily than words referring to more abstract things. Concreteness ratings were extracted from the MRC Psycholinguistic Database (Wilson, 1988). Of the Infant-CDI words, the 86 that were not found in the MRC database were rated by 7 research assistants, using the existing numbers for calibration, and the average of these was taken. The scale ranges from 100 (low) to 700 (high) but for computation was divided by 100 to range from 1 to 7.

Word category—This final predictor was a rough 3-way classification of the words by category. The categories and their counts in the complete dataset were: *noun* (227 words); *predicate* (essentially verbs and adjectives: 92); *closed-class* (45). The division was based on Bates et al. (1994), who also included a small “social words” category of nouns, which here were folded into the larger *noun* category. The categories are rough in part because using finer-grained categories would force more words into multiple categories. Even the division between noun and predicate is approximate because of the profusion of English words that are both nouns and verbs. In such cases words were assigned by hand based on judgments of sentences sampled from the corpus.

Note that although toddlers are known to learn a disproportionate number of words for objects, this does not imply that the nouns would have an advantage in the present models.

In estimating the effect of “noun” on the likelihood of knowing a word, nouns that a child did not know counted against the noun category just as much as nouns that the child did know counted for it. The fact that the CDI is, for good reasons, biased toward nouns (because children are too) means there are many listed nouns that most children do not know at 12 or 15 months. A consequence is that syntactic category is more important as a moderator of other effects than as a primary actor in the explanatory models, as will be shown below.

Results

Methodological preliminaries

The outcome measures in the dataset are a collection of ones and zeros: a yes or no for each child ($n=8$) X word ($n=364$) X age (12, 15 months) X response (understands, says) combination. Because so few words were reported as said at 12 months (a total of 93 words said, 63 of which were said by two children, and the remaining 30 spread out over four), this measure was not analyzed. We analyzed the “says” results and the “understands” results separately, because these outcomes are not predicted by the same factors in other research (Stokes, 2015).

The 15-month-olds’ production outcomes were analyzed using multilevel logistic regressions. The comprehension outcomes were analyzed using cumulative link mixed models (Agresti, 2002). These are ordinal regressions that are conceptually similar to performing a series of logistic regressions. This method provides a way to handle ordered outcomes with more than two values. The advantage of such models over, for example, performing separate logistic regressions at 12 months and 15 months, is that they respect the longitudinal nature of the dataset, with the same children and same words being evaluated two times. For each child and each word, three outcomes were considered: 00 (not knowing the word at either time point), 01 (not knowing the word at 12 months, but knowing it at 15), and 11 (knowing the word at both 12 and 15 months). The outcome 10 (knowing at 12 but not 15) was converted to 11; this is probably sometimes wrong, but conceptualizing word learning as a path from not understanding to understanding is what permitted treating word knowledge as an ordered factor.

The use of cumulative link modeling allows us to collapse the 12-month-olds’ and 15-month-olds’ results into one longitudinal analysis. A disadvantage of this approach is that it does not estimate whether 12-month-olds and 15-month-olds differ in which factors influence word learning the most. Analyses that attempt this are presented as Supplementary Materials, including separate regressions for 12 and 15 month olds, and an analysis that considers only *gains* in word knowledge from 12 to 15 months (i.e., comparing only 01 against 00).

In each of their analyses, B&S removed any word that was not understood (for comprehension analyses) or said (for production analyses) by at least one child of the eight. In our analyses we conservatively followed B&S’s procedure. Contrary to B&S, however, (a) if a word was listed as produced on the CDI, we assumed that the word was also understood; and (b) when a CDI was missing for a given child at a given age, we treated it as

missing data, whereas B&S used the next available CDI for that child (i.e., a CDI collected 3 months later). All together, these provisions reduced the set of word types being predicted from 357 to 339 words (“understands” analysis) and 215 words (“says” analysis).

Logistic regressions were computed in R (R Core Team, 2015, version 3.2.0) using the *glmer* function of the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) and the *clmm* function of the *ordinal* package (Christensen, 2015). Highly skewed variables (the frequencies) were log-transformed, and variables were centered (but not scaled, to keep the coefficients interpretable in terms of what they measure). In general, the initial models were as full as possible given the hypotheses, and models were simplified by removing variables that did not improve fits or that were not needed for the models to converge (where “full” means including separate slopes for different mother-child pairs, and as many theoretically reasonable interactions as the fitting could support). Variables that approached significance were retained in the models to help estimate their importance. Models were verified informally by graphical means and by re-running the final models 6 additional times each leaving out one mother-child pair, to be sure no conclusions depended on a particular child. These tests yielded results very similar to the complete models.

The large number of predictors we wished to evaluate required some constraint on the set of hypotheses tested, in that not all possible interactions could reasonably be included in the model. In the initial analyses, priority was granted to four factors: total word frequency, isolated-word frequency, duration, and the 3-way syntactic category division. Inclusion of the duration measure required that the dataset be restricted to the words for which acoustic data were available; for example, the analyses of word understanding that included the duration predictor included 2160 observations rather than 2768, and an average of 270 rather than 339 word types. When duration was not a relevant predictor, models using the larger dataset without this restriction are reported.

Word comprehension

Because one child (coded s2) was missing the 12-month CDI, the word comprehension analyses concerned seven children. The number of words each child had in each outcome category is given in Table 2, with children ordered by the number of words they already understood at 12 months. The table also presents the odds of a word being in the given outcome category, for each child. Odds are given because the analysis results reflect the impact of the predictors on the odds of the outcomes rather than their probabilities. Odds below one indicate that most words did not fall into that outcome category; odds above one indicate that most words did, for that child. For reference, note that odds of .667 correspond to a probability of 0.40, i.e. 0.40/0.60. Child s2 is not shown because the 12-month CDI was missing.

All seven of the children knew more nouns than words from other categories at both 12 and 15 months, but the *proportion* of the listed nouns known, relative to other categories, varied widely across children. For example, at 15 months child v2 was reported to know 47% of the nouns, but virtually all of the closed-class words. The totals and proportions are shown in Figure 1. These strong but divergent inter-child patterns were handled in the analyses by fitting child-specific slopes to the random effects of word type in the regression models.

The final model included total frequency, isolated-word frequency, concreteness, word category, and the interaction of word category with isolated-word frequency; in addition, MLU was retained though not significant ($p = 0.072$) to allow comparison of its effect size with the others (see Appendix). Median word duration ratio, neighbor counts, bigram probability or type frequency, and number of syllables were not retained. By far, the strongest, most consistent predictor was word frequency (Figure 2). An increase of one in log frequency increased the odds ratio of moving from not-knowing to knowing (at 12 or 15 months) 1.613 times. For example, child v1's odds of a word having the outcome 11 (understood already at 12 months, versus either of the other outcomes) were 0.648. Were a word to be heard more frequently, in the amount of an increase of one in its log frequency, we would expect the odds of outcome 11 to instead be $(0.648 * 1.613)$ or 1.045—a probability of 51.1% rather than 39.3%. An increase of one in log frequency corresponds to (for example) hearing a word 3 times (the overall median over all CDI words), to hearing it about 10 times. Coefficients, odds multipliers, interquartile ranges, and p-values for total frequency and the other significant predictors (and nonsignificant associated levels and interactions) are given in Table 3. Note that in the case of isolated predicate and isolated closed-class words, the coefficient given is the sum of the coefficient for nouns (the reference level) and the interaction coefficient for the case of the word being a predicate or closed-class. For the isolated-frequency interactions, the p-value is derived using the standard errors of the base case and the interaction terms and their covariance.

Brent & Siskind reported that total frequency was not a significant predictor. This non-effect (with the significant effect of isolated frequency) is readily replicable in the present dataset in a simple linear model if raw frequency counts, rather than their logarithm, are used as predictors. This implementational difference appears to account for the difference in our results. Because the Zipfian (Pareto, power-law) distribution of raw word frequencies leads them to be fitted poorly by regression models (whether linear or logistic), it is customary to use log frequencies instead in psychological and linguistic modeling (e.g., Baayen, 2001; Massaro et al., 1980; Schreuder & Baayen, 1997). A poorly-fitting model can obscure the clear relationship between frequency and learning; thus, in the present work, all frequencies were log-transformed.

Frequency in isolation had a significant effect, but the size of the effect varied by word type (Figure 3). For nouns, an increase of one in log isolated-frequency raised the odds of knowing a word by 1.76 times. For closed-class words, an increase of one in log isolated-frequency raised the odds by 1.98 times. These effects are quite strong. That said, most (85%) of the words were never heard in isolation at all; the 86th percentile corresponds to one instance, and the 92nd percentile corresponds to two. Strikingly, frequency in isolation had no measurable effect on the acquisition of predicates (an estimated coefficient of -0.017 , with standard deviation of about 0.188). For example, all mothers but one said the word *look* in isolation more than 25 times, yet only 3 children were said to know it.

Although we did not classify isolated words as utterance-final, they are, naturally, the last (and first) words in their utterance. Is the isolated-frequency effect actually an utterance-finality effect? This is a plausible hypothesis, given that mothers tend to elongate utterance-final words in speech to infants (Fernald & Mazzie, 1991), and given that one-year-olds

recognize utterance-final words more readily than utterance-medial words (Fernald, McRoberts, & Swingley, 2001). A prediction that follows is that utterance-final frequency would predict understanding better than utterance-nonfinal frequency. This turns out to be false: Their predictive power was very similar in a regression model equivalent to our final model (described above) but replacing total frequency with utterance-final, utterance-medial, and utterance-initial frequencies as predictors. The beta coefficients were 0.34 (std. err. 0.07) for final, 0.32 (0.08) for nonfinal (both $p < 0.0001$), and -0.02 (.10) for initial (ns). Thus, and somewhat surprisingly, utterance-final word position did not present evidence of playing a special role, and utterance-initial position showed no promise with the other word frequency predictors in the model.

Higher concreteness led to words being more likely to be understood, even with word category included in the analysis. An increase of one in concreteness (about 0.8 sd) raised the odds of knowing a word by 1.42 times.

The idea that more concrete words would be easier to learn has a long history in the study of word learning, traditionally in explaining why toddlers learn nouns in greater numbers than verbs (e.g., Gentner, 1982, who ties the basic idea of nominals' conceptual simplicity back to Aristotle). Here we find that the same is true even within categories: for example, children apparently found high-concreteness predicates like *kiss* and *hot* easier to learn than lower-concreteness predicates like *hurry* and *careful*.

In sum, analyses of word understanding at 12 and 15 months partially supported Brent & Siskind's original conclusions. Isolated-word frequency was a significant predictor, just as B&S argued, though this effect was moderated by word type and dwarfed by the effect of overall (log) frequency. Independently of frequency, concreteness was linked to understanding.

Word production at 15 months

One child (f1) was not reported to say any words, so all analyses considered only the remaining seven children. At 15 months, the number of words these children said ranged from 14 to 152, with a mean of 54 and *sd* of 48 (see Table 4). If including words that lacked acoustic duration data, the range was 16–175 (mean, 64). As in the comprehension dataset, children differed in their proportions of words said in each category (Figure 4).

The final model included total frequency, isolated-word frequency, and word category. MLU and word duration ratio were retained in the model though not significant ($p=.052$, $p=0.123$) to allow comparison of their effect size with the others (see Appendix). The coefficients and other statistics for the significant predictors are shown in Table 5.

Predictors that never improved model fit (or came close) included concreteness, bigram probability, bigram type frequency, neighbor counts, number of syllables, or any of the tested interactions.

Once again the word-frequency variables were the strongest predictors. An increase of one in log frequency increased the odds of saying a word by 1.32 times, and an increase of one in log isolated-frequency increased the odds 1.68 times. These effects are slightly smaller

than in the comprehension analysis, but here the effect of hearing a word in isolation was present across the three word types. These effects are illustrated in Figures 5 and 6.

Once again it did not seem that the isolated-word effect was actually a form of utterance-final advantage. As in the comprehension analysis the production model was re-run with the total-frequency term replaced by initial, medial, and final log frequencies as predictors. In this model, utterance-final frequency was not even significant ($p > 0.5$), though utterance-medial frequency was ($\beta = .35, p < .01$), and isolated frequency was quite strong ($\beta = .57, p = .0002$). Utterance-initial frequency was not a significant predictor ($p > .3$). Duration ratio was a marginal predictor in this analysis ($\beta = .34, p = .08$), but removing duration as a predictor did not improve the predictive power of utterance-final frequency. Thus, the isolated-word advantage did not reduce to a general advantage for utterance-final words.

When words tended to occur in long sentences, children were somewhat less likely to say them. The mean MLU, computed over moms and words (types), was 4.7 (sd across moms, 0.33; range 4.3–5.3). According to the regression model, an increase of one in median MLU would be expected to decrease the odds of the child saying the word by a factor of 0.89 times. For example, a child having .28 odds of saying a word on the CDI (the median in the present sample of 7; a probability of about 21.8%) would be predicted to say a word with (.28 * .89 \approx .25) odds, or about 20%, with an increase of one in median MLU. Density plots showing this effect for each child separately are given in Figure 7.

Children were also more likely to say words that were commonly spoken with long durations, but this effect was marginal in our main analysis. As described above, this measurement was computed relative both to the length of the word, and the speaker's typical rate. Many of the most-commonly elongated words were words for things or body parts, including *eye* (mean ratio 2.6), *shoe* (2.3), *bear* (2.1), or *nose* (2.0); but at the same time, many semantically similar words were not particularly elongated, such as *mouth* (1.1), *sheep* (1.3), or *pants* (1.5). An increase of 0.5 in the duration ratio (for example from the 35th percentile, 1.25, to the 79th percentile, 1.75), was estimated to increase the odds of saying a word by 1.16 times. In practice, this suggests that mothers would need to slow down their average speaking rate for a word quite substantially to have a large impact on their child's vocal production for that word at 15 months. Density plots showing the duration ratios for said and not-said words for each mother/child pair are given in Figure 8.

To summarize, the results again supported Brent & Siskind's (2001) claim that isolated-word frequency predicts word production in individual 15-month-olds. Other predictors were also important: total frequency, MLU (which makes words less likely to be said), and perhaps exaggerated duration (which makes words more likely to be said).

Maternal specificity

A special feature of the B&S dataset is the provision of both environment and outcome results for individual children. Is this link an important feature of our results, or are the measured linguistic environments interchangeable? One way to assess this is to run the models again, but using one mother's production data to predict word learning in other mothers' children. If parent-child pairs are aligned, as opposed to simply being samples

from a uniform language distribution, models using parent α to predict child β should fit poorly.

To this end, we generated all possible pairings of mothers' input data and children's outcomes, and compared the regression results for the true pairings (every mother with her own child) with the false pairings. If the relationship between a predictor and its outcome is child-specific, then we would expect that only a small number of regressions would (by chance) yield a stronger predictive value than the actual mother-child pairing. The results can be illustrated by considering the number of p values in the regressions that indicate a more reliable result than the true one.

In the *understands* analysis, the predictors were stronger for correct pairings than for random pairings for most variables. For total frequency, none of the 5039 random pairings exceeded the predictive reliability of the single correct pairing. For isolated frequency, only 3.1% did; for concreteness, none did; for MLU, 20.2% did; for word type *predicate*, 0.3%; word type *closed-class*, none. Thus, the only case where we failed to find considerable mom-child specificity was the case of MLU, which may not be surprising given that a portion of the MLU data is shared among the mothers. In the *says* analysis, true predictors again outperformed most of the randomized ones: total frequency, 2.9%; isolated frequency, 3.7%, MLU, 5.2%, duration ratio, 5.7%, word type *predicate*, 10.1%, word type *closed-class*, 4.8%.

These analysis indicated substantial maternal specificity. This being said, the precise specificity estimates given above involve a complex partitioning of variance among the variables—as illustrated by the fact that the predictive value of concreteness was greatest in the original “understands” analysis, even though the concreteness values were the same across children.

Another way to consider the issue of variability among mothers is to examine which lexical variables were most strongly correlated. Considering all pairwise correlations among mothers, log total frequency of words was strongly correlated, but log isolated frequency was much less strongly correlated. Correlations for the full set of important predictors that varied over mothers is given in Table 6.

The implication of this set of comparisons is that although there is some consistency in how mothers make words easier to learn (through higher frequency, shorter utterances, and so on), the specific words that are being emphasized in this way vary measurably from family to family. The statistical characteristics of speech in different households are different enough to have a strong impact on the likelihood that children will learn particular words.

Discussion

Our analysis found that samples of children's language environment predict those children's vocabulary outcomes. Overall word frequency was the most consistent and important predictor: when children hear a word many times, they are more likely to understand and say that word. In addition, children were more likely to understand and say words that appeared in one-word utterances, and words that otherwise tended to appear in shorter rather than

longer utterances. Words with more concrete denotations were understood more often, and words that tended to be exaggerated by mothers in their pronunciation (by lengthening, and probably other forms of hyperarticulation that typically go along with lengthening) were said more often (though this effect was not statistically significant). Here we consider each of these predictors, starting with isolated-word frequency, which was emphasized by the originators of the corpus.

Isolated-word frequency

Examining which words young children learn, given the language they hear, should tell us something important about how learning works. Our investigation began from Brent & Siskind's (2001) conclusion that young one-year-olds learn words most easily when those words are presented in isolation. The implication of this conclusion is that the problem of finding word-forms in their phonetic context is one of the most important limiting factors in the development of the early vocabulary.

What we found here, using the same sample, is that hearing words in isolation does indeed help 12- and 15-month-olds learn words (at least inasmuch as such conclusions can be drawn from correlational analysis). This effect does not reduce to any of a number of potential confounds: the potentially longer duration of words in isolation; their position at the end of utterances; their (rough) syntactic class; their high frequency in general; nor a global benefit from appearing in short (but not necessarily one-word) utterances. Either there is some lurking correlated feature of isolated words that we have not measured (i.e., a peculiarity that causes words to be said in isolation and also makes them easier to learn), or appearing in one-word utterances does aid in word learning.

Why is this? The usual explanation, as given above, concerns word segmentation: discovering the lexical units in a phonetically continuous, multiword stream of speech is difficult. This has been shown in numerous studies of infants and adults (e.g., Jusczyk, Houston, & Newsome, 1999; Mattys & Jusczyk, 2001; Seidl & Johnson, 2006; see Cutler, 2012, for a comprehensive overview). Another possibility is that words appearing in isolation are free of phonetic context effects that might lead infants to have only a vague impression of the words' sounds (a hypothesis that could in principle be tested by comparing words that are differentially affected by their most frequent contexts). It is also possible that infants' cognitive resources for attending to and encoding utterances are limited by quantity, and one word is simply easier to store in memory than two or more.

Overall frequency

Perhaps the least surprising of our results is that overall frequency is a strong predictor of word learning. We were able to replicate B&S's extraordinary failure to find a frequency effect by using raw frequency rather than log frequency, but the log frequency provides a better match to the data and to common practice in psycholinguistics and computational linguistics. The frequency effect itself does not resolve its cognitive mechanism, however, including the question of whether infants learn a little bit from each instance they encounter, or learn a lot from infrequent but very informative instances (Trueswell, Medina, Hafri, & Gleitman, 2013; Yurovsky & Frank, 2015; Yu & Smith, 2012).

Concreteness

The impact of concreteness, though somewhat less strong than that of frequency, is also consonant with prior descriptions of children's word learning. Children tend to learn many words for manipulable objects. Perhaps more interesting here is the fact that this effect, when found, was not restricted to objects. Nor was it driven by famously "invisible" predicates like *think* or *know* (e.g., Gleitman et al., 2005). High-concreteness words like *eat* and *hug* were marked as known more than words like *hurry* and *make*.

Curiously, the effect of concreteness was strong in the comprehension analyses, but not in production. In principle, this could have been due to the greater importance of MLU and duration in the production analyses, but concreteness is positively correlated with duration ($r = 0.37, p < .001$) and uncorrelated with MLU ($r = 0.03, ns$); further, replacing duration with concreteness in the model does not yield a significant role for concreteness. Thus it is unlikely that these other predictors muscled concreteness aside.

One explanation for this result is that concreteness is an important factor in children's discovery of word meaning, but once the word is understood, the barriers to speaking the word are not conceptual at all, but bear more on the child's phonological skills (in interaction with their parents' ability to understand their utterances). In our analysis the phonological variables like syllable number and phonotactic probability did not have significant predictive power, but this does not rule out such a possibility.

Another potential explanation is that concreteness in the comprehension analyses but not the production analyses actually tells more about the mothers than their children. Although observational assessments of the reliability of parent report on the CDI have shown significant correlations between measures, establishing the validity of parents' judgments of comprehension in infancy is challenging (e.g., Feldman et al., 2000; Fenson et al., 1994). Mothers may believe their children more likely to know words for more concrete concepts, producing a bias that underlies the predictive value of concreteness in our analyses. Such a bias would seem less likely to apply to predictors like phonetic duration, bigram probability, or MLU.

Median length of utterance

Median MLU (maternal utterance length in words) was negatively associated with word learning, though the effect fell shy of statistical significance in our final models. A significant MLU effect in the same direction was reported by Braginsky et al. (2016) and Roy et al. (2015). This effect could be due to the same word-identification difficulties that also support the learning of isolated words: if a word tends to appear in the context of many other words, it may be harder to extract and remember, or its meaning may be harder to identify. Alternatively, the short-MLU advantage might be due to children's difficulty in producing the longer utterances that high-MLU words appear to demand more frequently. This would explain why the MLU effect was only reliable in the production analysis. Given that the MLU effect did not interact with word type, it is not likely to be primarily due to the syntactic demands of each word; rather, some words tended to be used to express more complex ideas in longer sentences (such as *window* in *wanna go look out the window?*)

relative to others (such as *raisin* in *like some raisins?*). These potential mechanisms are quite different from one another, but we cannot weigh them definitively here.

Phonetic duration

15 month olds were somewhat more likely to say words that their mothers spoke with exaggerated duration, though this effect was not very consistent. Not a hint of this effect was present in the comprehension data at the same age. This difference is surprising, given that in laboratory tests of word understanding, 15-month-olds perform better when words are produced with greater duration, even utterance-medially (Fernald, McRoberts, & Herrera, n.d., Fernald et al., 2001). Perhaps exaggerated duration is useful for word recognition in one-year-olds but not materially beneficial for learning words. Another possibility is that exaggerated duration is useful for learning, but relatively few such instances are required and we did not capture them. In the case of children's word production, children may be disposed to say words whose importance is signaled by mothers through prosodic lengthening. Or hearing words with longer durations may lead children to have more accurate representations of words' form (or more confidence in their knowledge) leading to a greater chance of producing the word. A final possibility is that in fact mothers exaggerate words after children say them, i.e. that child productions cause greater maternal word duration, rather than the other way around (e.g., Gros-Louis, West, & King, 2014).

Phonotactics

The phonotactic variables did not have significant effects. Number of neighbors was never a significant contributor; nor was bigram probability (over types or over tokens). Other studies have, on the contrary, found effects of both neighborhood density and phonotactic probability. Storkel 2009, in a study of 1;4–2;6 year olds' CDIs, found that children learned low-probability CDI words at younger ages than high-probability words. Low-probability words may be less confusable with existing words—they “stick out” more (Han, Storkel, Lee, & Yoshinaga-Itano, 2015; Swingley & Aslin, 2007). On the other hand, words with many phonological neighbors appear to be learned at younger ages than words with few neighbors (Storkel, 2009). And words with many neighbors predominate in the productive lexicons of young children with relatively small vocabularies (Stokes, 2010; Stokes, Bleses, Basbøll, & Lambertsen, 2012) though Stokes (2015) found that this was not true of the words children know but do not say.

These somewhat contradictory patterns are puzzling: if high-probability words are harder to learn because they are not distinctive, large-neighborhood words should be harder to learn for precisely the same reason. Perhaps large-neighborhood words are harder to learn at first, because they are confusable, but once learned are easier to remember because neighbors reinforce each other in memory when they are heard and used (Stokes, 2015; Storkel, 2009). Here, the phonotactic predictors were computed from the Brent corpus but not linked from mother to child; thus, for these measures we do not have reason to take our null results on these predictors as challenging the prior analyses, some of which involved hundreds of (slightly older) children. What we may say with more confidence is that our other significant predictors are unlikely to have been statistical proxies for phonotactic variables.

Summary

To summarize, we found that infants are reported to know and say words more often when those words are frequent overall, and when they are frequently used in one-word utterances. Considering receptive vocabulary at 12 and 15 months, and productive vocabulary at 15 months, the predictors of concreteness, MLU, typical duration, and phonotactic probability all influenced word learning to varying degrees among these outcome measures. Higher concreteness (independently of word class) was linked to word understanding; lower MLU and greater durational exaggeration were linked to word production.

This study has several limitations. There were only 7 subjects in each analysis. There is no guarantee that the parents and children were typical of American English learners, and although the overall corpus is large by infant research standards, analyses like this one, which depend on details of specific words, are inevitably weakened by the data sparsity that comes from short sampling windows (relative to the child's daily experience) and from the skewed nature of linguistic frequency distributions. As a result the study (like any other) may have missed or mischaracterized some relationships between children's environment and their language outcomes. In addition, the outcome measure was based on parent report, which may suffer from bias of various kinds, particularly in estimation of children's receptive vocabulary. And of course, the degree to which our results can be generalized beyond American English remains to be seen (Braginsky et al., 2016).

In spite of these limitations, the study is one of a small number in which fine-grained details in individual parents' ordinary language behavior were related to very specific linguistic outcomes (e.g., Cristia, 2010; Roy et al., 2015). We found consistency across mothers in the properties of words that made them easier for their one-year-olds to learn, but the specific words exhibiting those properties varied from family to family. The fact that individual children's language environments were related to the words they knew in dyad-specific ways points to the value of quantifying individual children's language environments to predict linguistic outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH grant R01-HD049681 to D. Swingley. Portions of the research reported here were presented at the International Conference on Infant Studies in 2014. Some of the initial groundwork for the analyses was performed by Allison Britt, formerly an undergraduate at Penn, as her honors thesis. The authors thank Michael Brent and Jeff Siskind for making the corpus available, and our reviewers for some excellent and constructive suggestions.

References

- Agresti A. Categorical data analysis Hoboken, New Jersey: Wiley; 2002
- Baayen RH. Word frequency distributions Dordrecht, The Netherlands: Kluwer; 2001
- Bates D, , Maechler M, , Bolker B, , Walker S. lme4: Linear mixed-effects models using Eigen and S4 [Computer software manual] 2015 Retrieved from <http://CRAN.R-project.org/package=lme4> (R package version 1.1-7)

- Bates E, Marchman V, Thal D, Fenson L, Dale P, Reznick JS, ... Hartung J. Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*. 1994; 21:85–123. [PubMed: 8006096]
- Bergelson E, Swingley D. At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the USA*. 2012; 109:3253–3258. Retrieved from . DOI: 10.1073/pnas.1113380109 DOI: 10.1073/pnas.1113380109 [PubMed: 22331874]
- Bloom P. *How children learn the meanings of words* Cambridge, MA: MIT Press; 2000
- Braginsky M, , Yurovsky D, , Marchman VA, , Frank MC. From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*; 2016
- Brent MR, Cartwright TA. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*. 1996; 61:93–125. [PubMed: 8990969]
- Brent MR, Siskind JM. The role of exposure to isolated words in early vocabulary development. *Cognition*. 2001; 81:B33–B44. doi:/10.1016/s0010-0277(01)00122-6. [PubMed: 11376642]
- Cartmill EA, Armstrong BF, Gleitman LR, Goldin-Meadow S, Medina TN, Trueswell JC. Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*. 2013; 110:11278–11283. DOI: 10.1073/pnas.1309518110
- Cates CB, Weisleder A, Mendelsohn AL. Mitigating the effects of family poverty on early child development through parenting interventions in primary care. *Academic Pediatrics*. 2016; 16:S112–S120. DOI: 10.1016/j.acap.2015.12.015 [PubMed: 27044688]
- Christensen RHB. ordinal—regression models for ordinal data 2015 (R package version 2015.6–28. <http://www.cran.r-project.org/package=ordinal/>)
- Cristia A. Phonetic enhancement of sibilants in infant-directed speech. *Journal of the Acoustical Society of America*. 2010; 128:424–434. [PubMed: 20649236]
- Cutler A. *Native listening: Language experience and the recognition of spoken words* Cambridge, MA: MIT Press; 2012
- Feldman HM, Dollaghan CA, Campbell TF, Kurs-Lasky M, Janosky JE, Paradise JL. Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*. 2000; 71:310–322. DOI: 10.1111/1467-8624.00146 [PubMed: 10834466]
- Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. *Monographs of the Society for Research in Child Development*. 1994; 59(5) Serial Number 242. doi: 10.2307/1166093
- Fernald A, Marchman VA, Weisleder A. Ses differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*. 2013; 16:234–248. DOI: 10.1111/desc.12019 [PubMed: 23432833]
- Fernald A, Mazzie C. Prosody and focus in speech to infants and adult. *Developmental Psychology*. 1991; 27:209–221.
- Fernald A, , McRoberts G, , Herrera" C. Prosodic features and early word recognition. Paper presented at the 8th International Conference on Infant Studies; Miami, FL. 1992
- Fernald A, , McRoberts GW, , Swingley D. Infants' developing competence in recognizing and understanding words in fluent speech. In: Weissenborn J, , Hoehle B, editors *Approaches to bootstrapping: Phonological, lexical, syntactic, and neurophysiological aspects of early language acquisition* Vol. I. Amsterdam: Benjamins; 2001 97123
- Gentner D. *Language development* Vol. 2. Hillsdale, NJ: Erlbaum; 1982 Why nouns are learned before verbs: linguistic relativity versus natural partitioning; 301334
- Gleitman LR, Cassidy K, Nappa R, Papafragou A, Trueswell JC. Hard words. *Language Learning and Development*. 2005; 1:23–64.
- Gros-Louis J, West MJ, King AP. Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy*. 2014; 19:385–408. DOI: 10.1111/infa.12054
- Han MK, Storkel HL, Lee J, Yoshinaga-Itano C. The influence of word characteristics on the vocabulary of children with cochlear implants. *Journal of Deaf Studies and Deaf Education*. 2015; :242–251. DOI: 10.1093/deafed/env006 [PubMed: 25802320]

- Hart B, , Risley TR. Meaningful differences in the everyday experience of young American children Baltimore, MD: Paul H. Brooks Publishing; 1995
- Hoff E. How social contexts support and shape language development. *Developmental Review*. 2006; 26:55–88.
- Johnson EK, Jusczyk PW. Word segmentation by 8-month-olds: when speech cues count more than statistics. *Journal of Memory and Language*. 2001; 44:548–567.
- Jusczyk PW, Houston DM, Newsome M. The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*. 1999; 39:159–207. [PubMed: 10631011]
- Landau B, , Gleitman LR. *Language and experience: evidence from the blind child* Cambridge, MA: Harvard University Press; 1985
- MacWhinney B. *The CHILDES project: Tools for analyzing talk 3*. Hillsdale, NJ: Erlbaum; 2000
- Majorano M, Vihman MM, DePaolis RA. The relationship between infants' production experience and their processing of speech. *Language Learning and Development*. 2013; 10:179–204. DOI: 10.1080/15475441.2013.829740
- Markman EM. *Categorization and naming in children* Cambridge, MA: MIT Press; 1989
- Markman EM, Wachtel GF. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*. 1988; 20:121–157. doi:/10.1016/0010-0285(88)90017-5. [PubMed: 3365937]
- Massaro DW, , Taylor GA, , Venezky RL, , Jastrzembski JE, , Lucas PA. *Letter and word perception: The role of orthographic structure and visual processing in reading* Amsterdam: North-Holland; 1980
- Mattys SL, Jusczyk PW. Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*. 2001; 27:644–655. [PubMed: 11424651]
- McCune L, Vihman MM. Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research*. 2001; 44:670–684. DOI: 10.1044/1092-4388(2001/054)
- Nelson K. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*. 1973; 38:1–2.
- Newman RS, Rowe ML, Bernstein-Ratner N. Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development. *Journal of Child Language*. 2015; :1–16. DOI: 10.1017/S0305000915000446
- R Core TeamR: *A language and environment for statistical computing* [Computer software manual] Vienna, Austria: 2015 Retrieved from <http://www.R-project.org/> (R version 3.2.0)
- Rowe ML. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*. 2012; 83
- Roy BC, Frank MC, DeCamp P, Miller M, Roy D. Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*. 2015; 112:12663–12668. DOI: 10.1073/pnas.1419773112
- Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592. DOI: 10.1093/biomet/63.3.581
- Schreuder R, Baayen RH. How simplex complex words can be. *Journal of Memory and Language*. 1997; 37:118–139.
- Seidl A, Johnson EK. Infant word segmentation revisited: edge alignment facilitates target extraction. *Developmental Science*. 2006; 9:565–573. [PubMed: 17059453]
- Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, Samuelson L. Object name learning provides on-the-job training for attention. *Psychological Science*. 2002; 13:13–19. DOI: 10.1111/1467-9280.00403 [PubMed: 11892773]
- Song JY, Demuth K, Morgan JL. Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America*. 2010; 128:389–400. DOI: 10.1121/1.3419786 [PubMed: 20649233]
- Stokes SF. Neighbourhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research*. 2010; 53:670–683.
- Stokes SF. The impact of phonological neighborhood density on typical and atypical emerging lexicons. *Journal of Child Language*. 2015; 41:634–657. DOI: 10.1017/S030500091300010X

- Stokes SF, Bleses D, Basbøll H, Lambertsen C. Statistical learning in emerging lexicons: the case of Danish. *Journal of Speech, Language, and Hearing Research*. 2012; 55:1265–1273. DOI: 10.1044/1092-4388(2012/10-0291)
- Storkel HL. Learning new words: phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*. 2001; 44:1321–1337. doi:/10.1044/1092-4388(2001/103).
- Storkel HL. Developmental differences in the effects of phonological, lexical, and semantic variables on word learning by infants. *Journal of Child Language*. 2009; 36:291–321. DOI: 10.1017/S030500090800891X [PubMed: 18761757]
- Swingley D, Aslin RN. Lexical competition in young children's word learning. *Cognitive Psychology*. 2007; 54:99–132. DOI: 10.1016/j.cogpsych.2006.05.001 [PubMed: 17054932]
- Taine H. Acquisition of language by children. *Mind*. 1877; 2:252–259.
- Tincoff R, Jusczyk PW. Some beginnings of word comprehension in 6-month-olds. *Psychological Science*. 1999; 10:172–175.
- Tincoff R, Jusczyk PW. Six-month-olds comprehend words that refer to parts of the body. *Infancy*. 2012; 17:432–444. DOI: 10.1111/j.1532-7078.2011.00084.x
- Trueswell JC, Medina TN, Hafri A, Gleitman LR. Propose but verify: fast mapping meets cross-situational word learning. *Cognitive Psychology*. 2013; 66:126–156. [PubMed: 23142693]
- Vihman MM, Croft W. Phonological development: toward a 'radical' templatic phonology. *Linguistics*. 2007; 45:683–725. DOI: 10.1515/ling.2007.021
- Weisleder A, Fernald A. Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*. 2013; 24:2143–2152. [PubMed: 24022649]
- Whipple GM, editor *The twenty-fourth yearbook of the national society for the study of education* Bloomington, IL: Public School Publishing Company; 1925
- Wilson MD. The mrc psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers*. 1988; 20:6–11.
- Young SJ, , Evermann G, , Gales MJF, , Hain T, , Kershaw D, , Moore G, , ... Woodland PC. *The HTK book, version 3.4* Cambridge, UK: Cambridge University Engineering Department; 2006
- Yu C, Smith LB. Modeling cross-situational word-referent learning: prior questions. *Psychological Review*. 2012; 119:21–39. [PubMed: 22229490]
- Yuan J, Liberman MY. Speaker identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*. 2008; 123doi: 10.1121/1.2935783
- Yuan S, Fisher C. Really? she blicked the baby?: Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*. 2009; 20:619–626. [PubMed: 19476591]
- Yurovsky D, Frank MC. An integrative account of constraints on cross-situational learning. *Cognition*. 2015; 145:53–62. [PubMed: 26302052]

Appendix

The input data were in a format with one row per child per word. The outcome being predicted in each regression (specified to the left of the tilde (~) in the invoking *R* command) is a vector (column) from the set 00,01,11 for the comprehension analyses and 0,1 for the production analyses. Elements to the right of the tilde are predictors, where .c in the name indicates that the predictor was mean-centered. The reference level of the word category variable was *noun*; interactions represent departures from the reference case. The notation (1 + predictor | mom) means that mom is a random effect with intercept varying by mom and the slope of that effect allowed to vary according to predictor. Interactions are specified using a colon, as in `isol.freq:word.category`. In the linear regression analysis, `family="binomial"` and `control=glmerControl(optimizer="bobyqa")` were added, the latter because it specifies a search algorithm that seemed to converge in more analyses.

Fixed effects in cumulative link mixed regression model for understanding at 12 and 15 months

The final model was specified as follows:

```
clmm( understands ~ total.freq.log.c + isol.freq.log.c + median.MLU.c
concrete.c + word.category + isol.freq.log.c:word.category
+ (1+word.category|mom) + (1|word))
```

predictor	coef.	(SE coef.)	t	p
intercept 00–01	−0.3449	(0.6680)	−0.5163	0.6060
intercept 01–11	2.0802	(0.6702)	3.1040	* 0.0019
log total freq.c	0.4781	(0.0522)	9.1614	* 0.0000
log isolated freq.c	0.5654	(0.2141)	2.6408	* 0.0083
median MLU	−0.0580	(0.0323)	−1.7971	+ 0.0725
concreteness	0.3531	(0.1410)	2.5036	* 0.0124
class(closed)	−1.6105	(0.7785)	−2.0686	* 0.0387
class(predicate)	−0.2099	(0.4712)	−0.4454	0.6561
class(closed):isol.freq	0.1176	(0.3003)	0.3915	0.6955
class(predicate):isol.freq	−0.5824	(0.2697)	−2.1590	* 0.0310
Observations	2,373			
Log Likelihood	−1,915.8			

Fixed effects in regression model for saying words at 15 months

The final model was specified as follows:

```
glmer( says.15 ~ total.freq.log.c + isol.freq.log.c + median.MLU.c +
duration.c + word.category + (1+word.category|mom) + (1|word))
```

predictor	coef.	(SE coef.)	t	p
Intercept	−1.0595	(0.5858)	−1.8085	0.0708
log total frequency.c	0.2754	(0.0792)	3.4761	* 0.0005
log isolated freq.c	0.5197	(0.1492)	3.4823	* 0.0005
MLU.c	−0.1147	(0.0591)	−1.9393	+ 0.0527
duration ratio.c	0.2922	(0.1894)	1.5422	0.1233
class(closed)	−0.7239	(0.5284)	−1.3701	0.1709
class(predicate)	−1.5665	(0.5284)	−2.9644	* 0.0031
Observations	1,218			
Log Likelihood	−530.8			

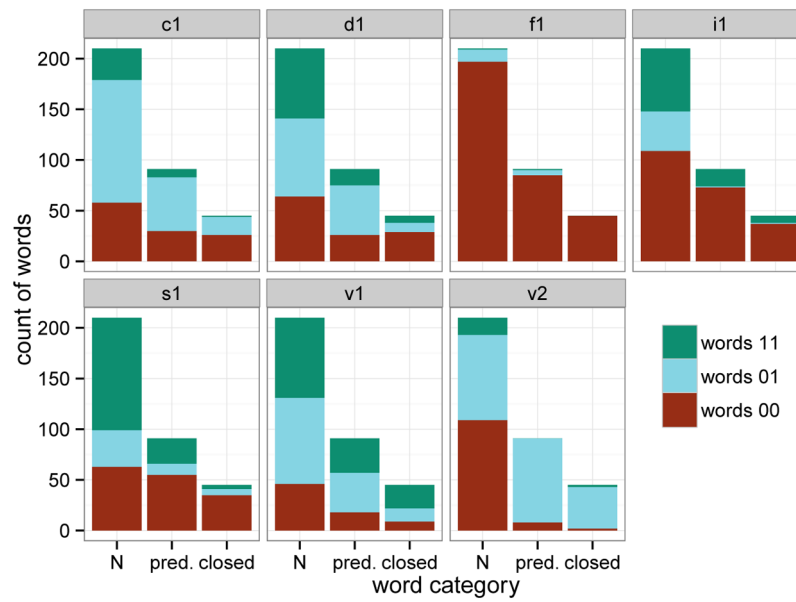


Figure 1.

Each child's count of words not known at 12 and 15 months (00), words not known at 12 but known at 15 (01), and words known already at 12 months (11), all split by word category (N = nouns, pred. = predicates, closed = closed-class words). Children tended to know more nouns, but proportionally speaking, often knew a greater fraction of the words in other categories. Each child is labeled by his or her code in the corpus files (c1, d1, f1. . .)

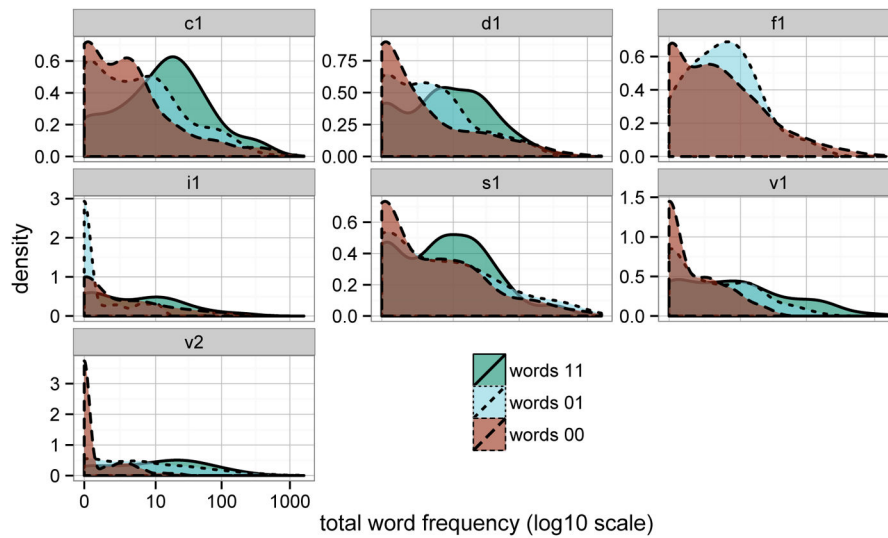


Figure 2.

For each child, distribution of total word frequency for words that child was reported to understand already at 12 months (words 11), to learn between 12 and 15 months (words 01) and to not understand at 15 months (words 00). The child's anonymous code-name (*c1*, *d1*, . . .) is given in the strip above each panel. Y-axes vary in each panel. For most children, the curve showing words 11 is to the right of the others (more frequent words are learned earliest) and the curve showing 01 is in the middle.

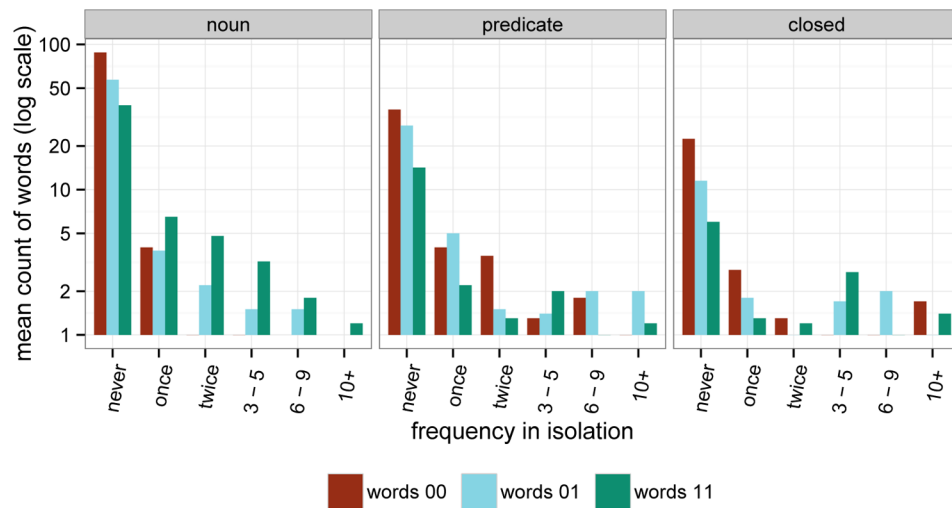


Figure 3.

Effects of words heard in isolation on children's understanding of the word at 12 months (words 11), learning between 12 and 15 months (words 01) and not understanding at 15 months (words 00). The y axis is on a log scale. At 12 and 15 months, children hearing nouns and closed-class words in isolation tended to learn those words. The isolated-word advantage was not evident for predicates.

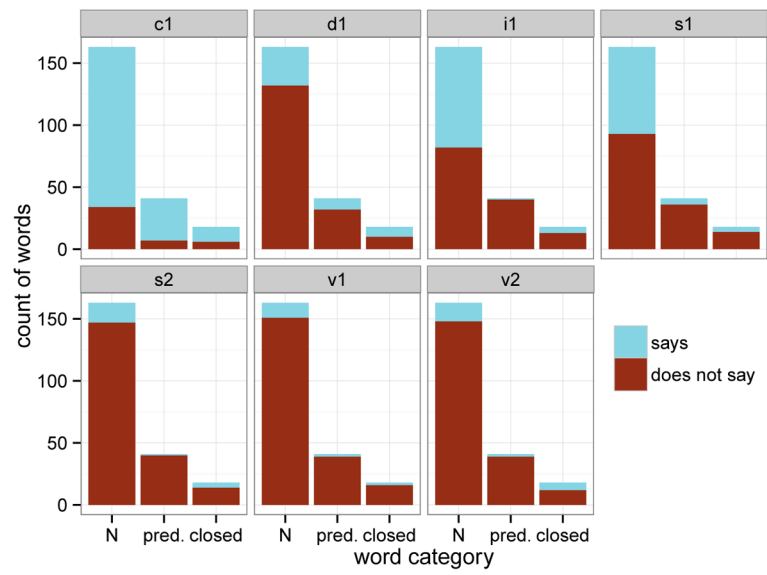


Figure 4.

At 15 months, each child's count of words said (light bars) and not said (dark bars) in each word category. *N* = noun; *pred.* = predicate; *closed* = closed-class.

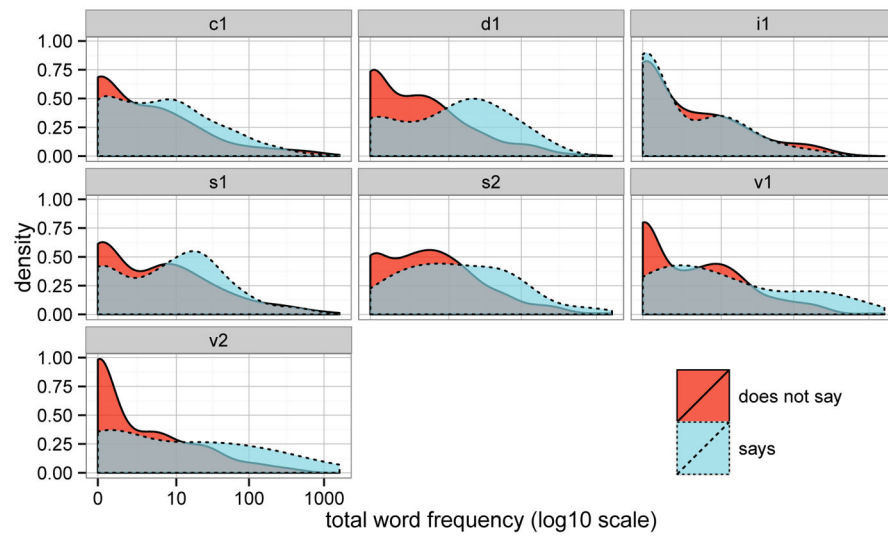


Figure 5. The distribution of total word frequency for the words that each 15-month-old was reported to say (in light blue) and not to say (in darker red). The child's anonymous code-name (*c1*, *d1*, . . .) from the corpus is given in the strip above each panel. The area under each curve is one.

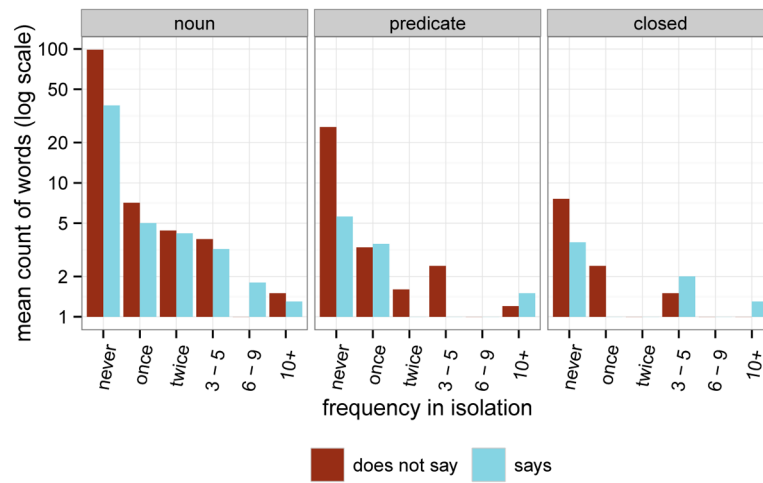


Figure 6. 15-month-olds' word production as a function of the words' frequency in isolation, and their word type. The isolated-word effect was not moderated by word type. The y-axis is plotted on a log scale, with tick labels giving mean counts over mother-child dyads.

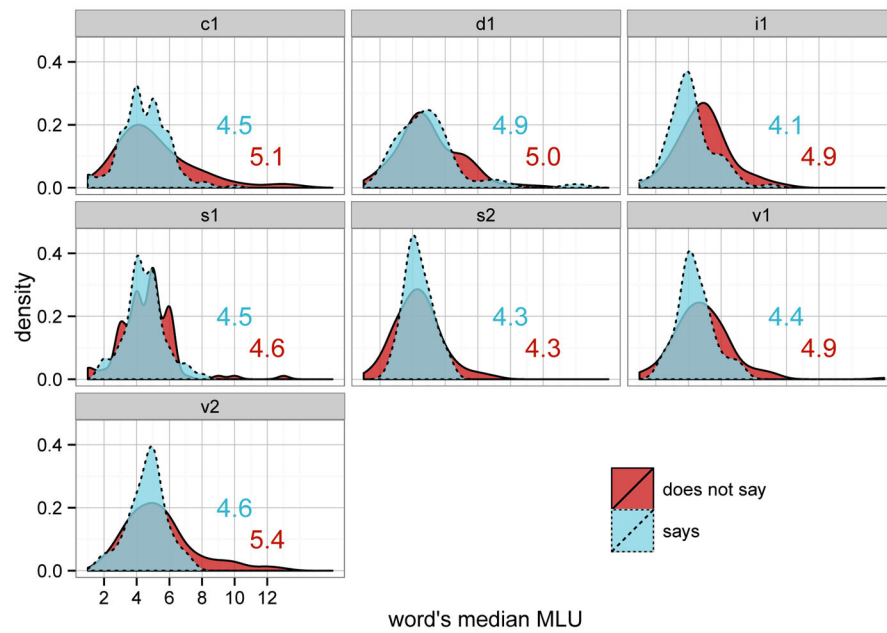


Figure 7.

15-month-olds' word production as a function of the words' median MLU, for each dyad. Light blue (dotted line) shows words the child said, darker red (solid line) shows words not said. The area under each curve is one. The mean MLU-median of words said (upper left) and not said (lower right), for each child, is given as numbers on each panel. Each child is labeled with his code in the Brent corpus (gray strips above each panel). All values are from each mother's corpus alone, with no imputation for missing data. The plot shows that words higher in MLU were somewhat less likely to be said by children.

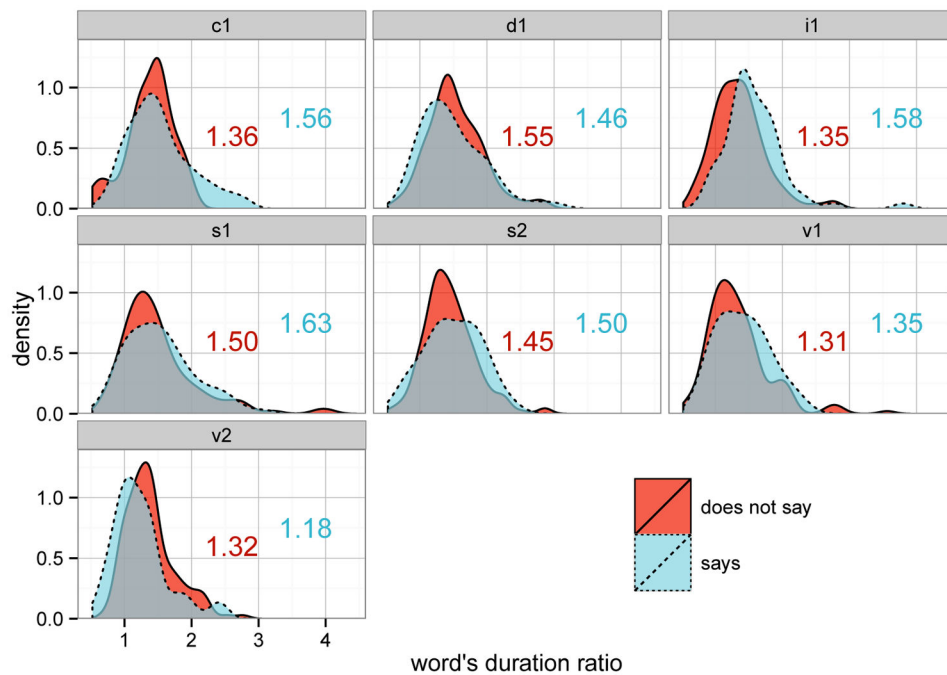


Figure 8. 15-month-olds' word production as a function of the words' duration ratio, i.e. how much longer the word's average realization was relative to the sum of the average durations of the word's component sounds. Light blue (dotted line) shows words the child said, darker red (solid line) shows words not said. The area under each curve is one. The mean duration ratio of words said (upper right) and not said (lower left), for each child, is given as numbers on each panel. Each child is labeled with his code in the brent corpus (gray strips above each panel). The plot shows that for most of the children, words that mothers said with exaggerated duration were more likely to be said by their children.

Table 1

Maternal word use for each child.

outcome	child	c1	d1	f1	i1	s1	s2	v1	v2	mean
count of word types		1792	1875	1878	1915	1566	2047	1872	1581	1816
count of word tokens		35780	34436	32397	39254	41408	46995	37178	26311	36720
type:token ratio		.050	.054	.058	.049	.038	.044	.050	.060	.050

Note. For each of the 8 children, the number of words that the mother said. Types were based on the orthographic rendering in the corpus after the cleaning-up steps reported in the text.

Table 2

How many words each child understood at each age

outcome child	f1	v2	c1	i1	d1	v1	s1
11: learned before 12 months	2	19	40	86	92	136	140
01: learned between 12 and 15	17	208	192	41	135	137	53
00: not understood by 15 mos.	327	119	114	219	119	73	153
odds of 11	0.0058	0.0581	0.1307	0.3308	0.3622	0.6476	0.6796
odds of 01	0.0517	1.5073	1.2467	0.1344	0.6398	0.6555	0.1809
odds of 00	17.2116	0.5242	0.4914	1.7244	0.5242	0.2674	0.7928

Note. For each of the seven children, the number of words that the child understood already at 12 months (first row), did not understand at 12 but understood at 15 (second row), or did not understand at either age (third row). Counts are expressed as *odds* of being in the given category (11, 01, 00) versus not being in it; e.g., for child s1, the odds of a word being in 11 (understood at 12 months) are ((140/346)/(1-(140/346))), or 0.6796. The total number of words to be learned is 346 for each child.

Table 3

Regression coefficients and descriptive statistics of significant predictors in the word-understanding analysis.

predictor	coef	exp(coef)	IQR	90-10R	p value
total frequency,c	0.4781	1.6129	2.56	3.85	.0000
isolated freq. (nouns)	0.5654	1.7602	0.00	0.69	.0083
isolated freq. (closed)	0.6830	1.9798	0.00	1.39	.0022
isolated freq. (pred.)	-0.0170	0.9831	0.00	0.69	> .9
MLU.c	-0.0580	0.9437	2.00	4.00	.0725
concreteness.c	0.3531	1.4235	2.00	3.17	.0124
class(closed)	-1.6105	0.1998	na	na	.0387
class(predicate)	-0.2099	0.8107	na	na	> .6

Note. *Coef* refers to the estimated beta coefficient from the ordinal regression model. *Exp(coef)* provides the number by which the odds of moving from 00 to 10 or 10 to 11 should be multiplied given an increase of one in the predictor's value. *IQR* (interquartile range) is the difference in value between the 75th and 25th percentiles for values of the numerical predictors. *90-10R* is like the IQR but uses the 90th and 10th percentiles. These give a sense of how many "increases of one" of the predictor's value are actually available in the range of the data. The IQR of isolated word frequency is zero because more than 75% of CDI words never occur in isolation.

Table 4

How many words each child was reported to say at 15 months

outcome child	cl	dl	il	sl	s2	v1	v2
not saying	46	168	132	139	194	199	192
saying	169	47	83	76	21	16	23
odds	3.674	0.280	0.629	0.547	0.108	0.080	0.120

Note. For each of the 7 children, the number of words that the child was reported to not yet say (first row), to say (second row), and the odds of saying the word (third row). Child fl is not shown because the CDI indicated no words said.

Table 5

Regression coefficients and descriptive statistics of significant predictors in the word-saying analysis.

predictor	coef	exp(coef)	IQR	90-10R	p value
total frequency.c	0.2754	1.3171	2.77	4.01	.0005
isolated freq.c	0.5197	1.6815	0.00	1.10	.0005
MLU.c	-0.1147	0.8917	1.00	3.00	.0527
duration ratio.c	0.2922	1.3393	0.53	2.42	.1233
class(closed)	-0.7239	0.4848	na	na	.1709
class(pred.)	-1.5665	0.2088	na	na	.0031

Note. *Coef* refers to the estimated beta coefficient. *Exp(coef)* provides the number by which the odds of saying a word should be multiplied given an increase of one in the predictor's value. *IQR* is the difference in value between the 75th and 25th percentiles for values of the numerical predictors. 90-10R is like the IQR but uses the 90th and 10th percentiles.

Table 6

Correlations of 4 predictors between mothers.

predictor	mean	min	max
log total frequency	0.61	0.52	0.70
log isolated freq.	0.29	0.12	0.39
median MLU	0.31	0.18	0.45
duration ratio	0.55	0.38	0.66

Note. Pairwise Spearman (rank) correlations between mothers for each of four continuous predictors are shown. There were 8 mothers and therefore 28 pairwise correlations for each predictor.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript