J. Child Lang. **35** (2008), 515–531.
doi:10.1017/S0305000907008641 Printed in the United Kingdom

Does frequency count? Parental input and the acquisition of vocabulary*

JUDITH C. GOODMAN
University of Missouri-Columbia

PHILIP S. DALE
University of New Mexico

AND

PING LI
Pennsylvania State University

(*Received* 1 *December* 2006. *Revised* 27 *July* 2007)

ABSTRACT

Studies examining factors that influence when words are learned typically investigate one lexical category or a small set of words. We provide the first evaluation of the relation between input frequency and age of acquisition for a large sample of words. The MacArthur-Bates Communicative Development Inventory provides norming data on age of acquisition for 562 individual words collected from the parents of children aged 0;8 to 2;6. The CHILDES database provides estimates of frequency with which parents use these words with their children (age: 0;7–7;5; mean age: 36 months). For production, across all words higher parental frequency is associated with later acquisition. Within lexical categories, however, higher frequency is related to earlier acquisition. For comprehension, parental frequency correlates significantly with the age of acquisition only for common nouns. Frequency effects change with development. Thus, frequency impacts vocabulary acquisition in a complex interaction with category, modality and developmental stage.

[*] Preparation of this article was made possible in part by a grant from the National Science Foundation (BCS-0131829) to PL. We thank Shihfen Tu for the initial calculations of age of acquisition from the CDI database and Janet Patterson for her helpful comments on the manuscript. We are grateful to the many child language researchers who contributed data to the CHILDES database that made possible the estimates of input frequency utilized here. Address for correspondence: Judith C. Goodman, Department of Communication Science and Disorders, 303 Lewis Hall, University of Missouri-Columbia, Columbia, MO 64111. Email: goodmanjc@health.missouri.edu

The acquisition of a large and diverse vocabulary is one of the major achievements of early childhood. The broad outlines of this accomplishment for English are well known (Bates, Dale & Thal, 1995): children begin to produce their first words, on average, at about 1;0. Children initially add words to their productive vocabularies slowly, increasing by only a few words each month. After about six months (when they produce about 50–100 words), their rate of vocabulary acquisition often increases sharply. This increased rate of lexical acquisition is often referred to as the 'vocabulary spurt' or 'naming explosion' (Dromi, 1987; Goldfield & Reznick, 1990; Li, Zhao & MacWhinney, 2007; Nelson, 1973). By 2;6, the median vocabulary size is about 570 words (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007); estimates of vocabulary size at the beginning of the school years often exceed 10,000 words (Carey, 1978). Less research has been devoted to studying lexical comprehension in very young children. It is widely agreed, however, that children start to comprehend words prior to production, and that comprehension vocabularies are even larger than production vocabularies (Benedict, 1979; Clark & Hecht, 1983). While this overall description has been well documented, it remains unclear why children learn particular words when they do.

The selectivity of early vocabulary is as remarkable as its quantitative growth. Very young children are exposed to millions of word tokens in adult speech, including thousands – probably tens of thousands – of different words (Hart & Risley, 1995; Weizman & Snow, 2001). From this 'sea of words', children learn notably similar early vocabularies (Bates *et al.*, 1995). For a majority of English-learning children, the earliest words refer largely to objects, although some personal–social and relational words are also present. When children produce about 200 words, they begin to add verbs and adjectives in greater proportions than earlier. Closed-class words begin to become frequent after 400 words.

Although a diverse range of theories has been proposed to explain both the number and categories of words learned (Golinkoff *et al.*, 2000), most appear to assume a positive role for frequency. That is, all other things being equal, the more often a specific word is heard, the earlier it should be learned. To be sure, some theories posit that the effect of exposure will depend on the interactive context, and therefore some exposures to a word will be more productive for learning than others. For example, Tomasello (2003) has emphasized the role of exposure during joint attention, and in situations where an adult's intention can be inferred, for lexical learning. Patterson (2002) provided evidence that exposure during book reading was more related to vocabulary growth than exposure from television. Tardif, Shatz & Naigles (1997) found that, along with frequency, utterance position and morphological variation affect the relative rate of acquisition of nouns and verbs in different languages. Similarly, Naigles & Hoff-Ginsberg (1998)

report that total frequency, frequency of occurrence in utterance-final position and occurrence in a greater range of syntactic frames all contribute to the order of acquisition of 25 verbs. Although these and other arguments suggest that the effect of frequency is not simply linear, there appears to be a general theoretical consensus on the positive effect of frequency, that the greater the frequency with which a word is produced in speech directed to children, the earlier it will be learned.

Most of the research conducted to date, however, provides only indirect forms of evidence for evaluation of the effect of frequency on vocabulary acquisition. First, some studies infer an effect for frequency on individual words based on overall vocabulary size. It is well established that parents who provide more input overall have children whose early vocabulary grows more quickly (Hart & Risley, 1995; Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991; Weizman & Snow, 2001). Second, training studies have manipulated exposure frequency on small sets of novel words. Schwartz & Terrell (1983) varied the frequency of novel words presented to one-year-old children. They found that frequently presented words were more likely to be learned than infrequently presented words. More frequent presentations may facilitate segmentation of words from fluent speech and provide a wider variety of extralinguistic contexts for inferring word meanings. These frequency effects are observed not only for normal monolingual children, but also in children with language disorders and second-language learners. For example, Rice, Oetting, Marquis, Bode & Pae (1994) controlled the number of presentations of novel words that occurred in stories, and found that ten presentations per word resulted in better word learning and retention than three presentations for children with specific language impairment (SLI), and Wang & Koda (2005) found that word frequency affected naming performance for Chinese and Korean college students learning to read English as a second language. Third, several studies have suggested that frequency affects which categories of words children learn. As noted above, American children use a large proportion of nouns in their early vocabularies (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994; Gentner, 1982). However, child-directed speech (CDS) in some languages, such as Korean and Chinese, uses verbs more frequently than does English CDS, and children's early vocabularies in those languages include a higher proportion of verbs than those of American children. This finding suggests that the frequency with which children hear words of particular syntactic categories affects the composition of their lexicons (Choi & Gopnik, 1995; Gopnik & Choi, 1990; Tardif, Gelman & Xu, 1999). Fourth, Harris, Barrett, Jones & Brookes (1988) and Barrett, Harris & Chasin (1991) found a strong relation between linguistic input and children's first word uses. They examined whether children's early words were context-bound, nominals or non-nominals. They found that the ways

in which children first used words were strongly tied to the frequency and patterns of use of these words by their mothers. For example, a mother who frequently used the word *there* in the context of her child handing her a toy was likely to have a child who initially used the word *there* when he was handing his mother a toy. Similarly, Brent & Siskind (2001) reported that a substantial fraction of the words infants aged 1;2 produce are words that mothers earlier spoke in isolation and that the frequency with which a child heard a word in isolation predicted lexical acquisition.

Although these studies suggest that input frequency is an important factor in vocabulary acquisition, their limitations leave open several questions. First, most studies have examined the developing lexicon at the level of syntactic–semantic categories or even more broadly, with respect to total size. That is, they have correlated the amount of mothers' speech with children's overall vocabulary size or correlated the frequency of syntactic categories in the input with the frequency of those categories in children's vocabularies. They have not, however, compared the frequency of individual words in mothers' speech with acquisition data for those same words for a relatively large lexicon. Thus, although they have demonstrated that mothers who talk more have children who know more words, and that parents in linguistic communities that produce more verbs have children who produce relatively more verbs, they cannot explain why particular words are learned early and others later. Second, those few empirical studies that have examined the acquisition of particular words have generally been limited to only a few novel words, as in training studies, or to a single lexical category. For example, Naigles & Hoff-Ginsberg (1998) examined the effects of verb frequency, as well as other variables mentioned above, and found that all were related to age of acquisition. Blackwell (2005) examined how frequency influenced the acquisition of words in the category of adjectives. She observed a correlation between the frequency of individual adjectives in maternal speech to Brown's (1973) subjects Sarah and Adam and the age of acquisition of those adjectives by the children (other characteristics of adjectives, such as syntactic diversity, also contributed to the prediction). While this is an important and relevant result, it and other studies at present do not permit assessment of the effects of a single variable such as frequency across the range of grammatical categories. Third, most studies of lexical acquisition have examined only children's productive vocabularies (but see Schwartz & Terrell, 1983) and thus cannot evaluate the role of input frequency on comprehension vocabularies. Fourth, estimates of the frequency of individual word use have generally been based on frequency in adult-directed written language (Kucera & Francis, 1967) or written language directed toward older children (Thorndike & Lorge, 1944) rather than in child-directed oral speech, except for a very few studies of restricted semantic subsets (e.g. Blackwell, 2005; Naigles & Hoff-Ginsberg, 1998).

Estimates of frequency based on written language directed to older children and adults are likely to be quite different from the actual frequencies in the speech to which young children just learning language are exposed, and therefore they do not actually capture the role of frequency for children's vocabulary acquisition.

Taken together, these limitations mean that the hypothesis that higher word frequency leads to earlier word learning has never been directly tested. There is the hint that it is true based on results for overall vocabulary size and the acquisition of single categories of words, but the way studies have been conducted prevent us from knowing just how far to take this hypothesis. In its strongest form, this hypothesis would predict that, across lexical categories, the more common the word the earlier children learn it. But this strong version must not be correct, because it is well established that closed-class words occur most frequently in adult speech but are not common in children's early vocabularies. It is important to know how important frequency truly is in acquisition. Although other variables are also important in determining when a child learns a word, a better understanding of the role of frequency in word learning is needed before we can study how it might interact with other variables to affect lexical learning. To do this, a comparison of the input frequency and age of acquisition of a large number of words from the full range of lexical categories is needed.

The goal of the present study is to investigate the relation between word frequency and children's early lexical acquisition. Unlike previous studies, it looks at a large number of words from the full range of lexical categories seen in early language. Further, it looks at the role that input frequency, based on child-directed speech, plays in word learning both across and within lexical categories. Finally, it examines the role of frequency in the development of comprehension vocabularies as well as production vocabularies.

These methodological advances are made possible by the availability of two large datasets. The first is the norming database of the MacArthur-Bates Communicative Development Inventories (CDI; Fenson *et al.*, 2007), a pair of parent-report questionnaires for the assessment of young children's communication, including vocabulary. Estimates of the age of acquisition of specific words are derived from this database. The second dataset is drawn from the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). The CHILDES system includes a database of transcripts of children's language production which have been made available for public use (see http://childes.psy.cmu.edu/). We used the English-language transcripts of parents speaking with toddler and preschool-age children in this database to establish the input frequency of individual words included on the CDI.

The research strategy followed here, which utilizes information on input frequency and age of acquisition from independent samples, has three

advantages over the more obvious one of drawing both input and age of acquisition measures from the same dyads. First, studies correlating children's acquisition of words with frequency in their own mothers' input are necessarily restricted to either a small number of words or a very small number of mother–child dyads, limiting our ability to generalize their findings. The present study uses a large database of parental speech to young children to establish word frequency and another large database to assess the age of acquisition of particular words. Thus the measures are likely to have increased reliability and validity. Second, the present strategy provides a very conservative test of input frequency. There might be a substantial correlation between frequency and age of acquisition, but frequency might vary greatly across dyads and, therefore, no overall correlation obtained. If in fact we find children's age of acquisition to be significantly correlated with input frequency from a different group of parents, then the effect is quite general. Third, this strategy eliminates a potential confound which could produce a spurious effect. Individual mothers might be particularly interested in (and, therefore, more frequent users of) specific categories such as object names, actions or words about emotions, due to their genetic make-up. This make-up is shared with their children, who might also demonstrate emphasis on selected vocabulary categories for the same reason. For example, it is possible that children with a referential style of language acquisition (Nelson, 1973) may have parents who also use a greater proportion of nouns because of a genetic predisposition that they pass on, in contrast to dyads who share a greater interest in social interaction and relevant language forms, i.e. an expressive style. Although this is, at present, only speculation, other verbal behaviors show a genetic load. By using independent samples, we avoid correlations which would represent only shared genetic endowment.

METHOD

Determination of age of acquisition of individual words

The CDI norming dataset may be used to establish norms for individual words (Dale & Fenson, 1996) by determining the percentage of children who know each of the words. The CDI: Words and Gestures (CDI:WG), for children who are aged 0;8 to 1;4, includes 396 words. Parents were presented with a list of words organized by categories and asked to fill in one bubble if their child understood a word and another bubble if their child both understood and said the word. The CDI: Words and Sentences (CDI:WS), for children who are aged 1;4 to 2;6, includes 680 words, also organized by categories. Parents filled in a bubble next to each word if their child produced that word. Children's comprehension is no longer assessed, because it has grown to a point where parents cannot reliably keep track of

it (Fenson *et al.*, 1994). Both forms include other items as well (e.g. gestures on the CDI:WG and grammar on the CDI:WS), but for the present study we have used only the lexical data. We defined the age of acquisition of a word in comprehension or production as the first month in which 50% of the children in the norming study were reported to comprehend or produce it, respectively.

Determination of parental input frequency

We estimated parental frequency by searching all parental/caregivers' speech transcripts in the CHILDES database for every use of the items on the MacArthur CDI forms (Li, Burgess & Lund, 2000). We used the following 28 CHILDES corpora: Bates, Belfast, Bernstein, Bliss, Bloom70, Bloom73, Brown, Clark, Cornell, Demetras1, Demetras2, Fletcher, Gathercole, Hall, Higginson, Howe, Kuczaj, Macboys (MacWhinney), Peters, Post, Sachs, Snow, Suppes, Valian, Vanhouton, Vankleeck, Warren, Wells. The children in these transcripts ranged in age from 0;7.23 to 7;5, with a mean of 36 months; most (about 75%) of the transcripts were under 4;0.

The total number of lexical items in this corpus is approximately 3·8 million word tokens. We excluded words from analysis in the present study if: (a) fewer than 50% of children acquired them by age 2;6 according to the CDI norming dataset; (b) if they were animal sounds or sound effects; (c) if they were multiword forms (e.g. *on top of*) on the CDI; (d) if they were proper names (e.g. a pet's or babysitter's name) on the CDI; or (e) if they were words with two common uses, such as *swing* (action vs. place) and *clean* (description vs. action), because the two meanings could not be distinguished computationally in the input. We also intended to exclude words if parents did not produce them in any of the CHILDES transcripts, because, like the words in category (a), they would not provide data on both variables for correlations. However, this category did not occur for any words which met our other criteria. These criteria resulted in age of acquisition and input frequency estimates for 562 of the words on the CDI. Each word was assigned to one of six lexical categories: common nouns, people words, verbs, adjectives, closed class and others. We chose these categories because they correspond to the groupings of the CDI, and have been widely used in other research. Table 1 shows examples of each category along with the number of words in that category. Note that 'common nouns' refer to objects and substances; other nouns such as those labeling events (e.g. *lunch*) and locations (e.g. *park*) are in the category 'others'.

For comparison with input frequency based on child-directed speech in CHILDES, we also examined frequency data from the Kucera–Francis (1967) and Thorndike–Lorge (1944) norms, which are based on written

TABLE 1. *Categories and numbers of words analyzed in the present study, based on parent report from the CDI and parental production in CHILDES transcripts*

Lexical category	Example words	Number of words
Common nouns	ball, frog, juice	256
People words	doctor, girl, mommy	21
Verbs	bite, hug, take	90
Adjectives	big, happy, tired	55
Closed class	that, in, some	68
Others	please, lunch, park	72
Total		562

language produced by adults for adults, and on materials prepared for older children, respectively.

RESULTS

Because frequency of occurrence is highly skewed for the 562 words in our dataset, log frequencies were used in all analyses; raw frequencies will be reported in examples. Pearson product-moment correlations were computed to establish whether the age of acquisition of specific words was related to the frequency with which parents use those words. These correlations were computed for the set of words as a whole and for the individual lexical categories. In addition, correlations were computed to examine whether age of acquisition was related to frequency of the words in adult-to-adult language as assessed by the common frequency scales in the psycholinguistic literature. Note that the overall hypothesis of the present study predicts a negative correlation between input frequency and age of acquisition. For ease of interpretation, we report correlations between the negative of log input frequency and age of acquisition. This correlation is hypothesized to be positive if higher frequency leads to earlier acquisition.

We asked first whether parental frequency is correlated with age of acquisition for specific words, defined by production, by examining the entire set without subdividing it into lexical categories. The correlation across 562 words was negative ($r = -0{\cdot}068$; $p = 0{\cdot}055$). Within the present sample of words, the more frequent a word in the parents' usage the later it is learned. This finding may seem surprising at first glance, since most theories would predict that the more frequent a word is, the earlier a child would learn it. However, when one considers which words are most frequent, the effect makes sense: closed-class words were produced most frequently, on average, in parental input (mean frequency = 16,116 in the total corpus; that is, individual closed-class words, on average, occurred

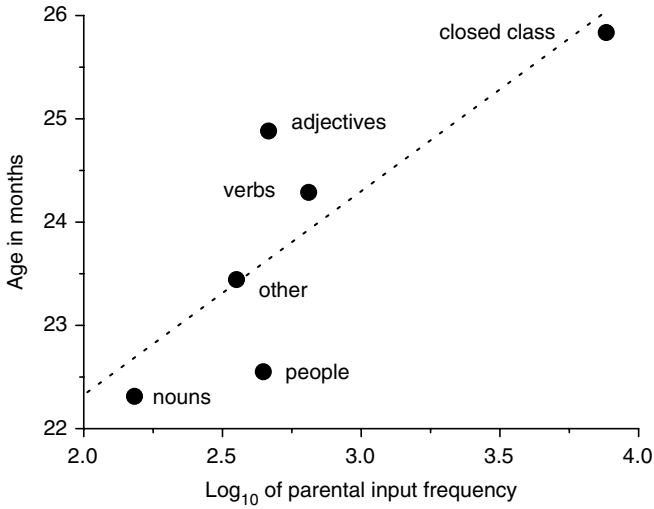

Fig. 1. Mean parental frequency and mean age of acquisition for six lexical categories.

16,116 times each across all samples in the corpus), but tend to be learned the latest, at least within the 30-month time window for our study. In contrast, individual nouns were produced least frequently, on average, in parental input (mean frequency = 309), but tend to be learned the earliest. Figure 1 displays the relation between parental frequency and children's learning, based on means for each category.

We then looked at the relation between parental frequency and the age of acquisition of words within lexical categories. These analyses were performed separately for production and comprehension data. Turning first to the production data (using both the CDI:WG and the CDI:WS data), we obtained the expected positive correlation. That is, for each category, the higher the frequency of a word within a category, the earlier that word is acquired. The correlations are shown in Table 2.

Importantly, the age of acquisition of specific words by children based on the norming study is much more strongly related to parental frequency as estimated using the CHILDES database than it is to the frequency of the same words in commonly used measures of word frequency that are based on either adult-to-adult or older children's written discourse. Table 2 also shows the correlations for the relation between age of acquisition and the Kucera–Francis and Thorndike–Lorge frequency norms. Although the correlations between age of acquisition and frequency reach statistical significance for all categories when frequency estimates are based on parental speech, the correlations are much weaker, and significant only for common nouns, when adult frequency norms are used. These substantial

TABLE 2. *Correlations between age of acquisition in production and selected estimates of word frequency*

Category	*N*	*r* (parental frequency)	*r* (Kucera–Francis)	*r* (Thorndike–Lorge)
Common nouns	256	0·55**	0·21**	0·23**
People words	21	0·52*	−0·05	−0·04
Verbs	90	0·22*	0·00	0·05
Adjectives	55	0·28*	−0·15	−0·11
Closed class	68	0·24*	0·00	0·01
Others	72	0·34**	0·11	0·16

* $p<0·05$, ** $p<0·01$.

correlations demonstrate that even though our estimates of input frequency are derived from a group of parents that is independent of the children from whom our age of acquisition estimates are derived, the frequency of words directed to young children is sufficiently uniform, and sufficiently different from the frequency of words directed to adults, to contribute to our understanding of the factors underlying vocabulary acquisition.

It is possible that the analyses above somewhat overestimate the role of frequency, as we have excluded a category of words that is characterized by discrepancies between age of acquisition and input frequency, namely, words that are not acquired by age 2;6 but that are reasonably frequent in the input. There were 31 words that occurred at least 300 times in the input (this is the average frequency for nouns, which have the lowest mean frequency of any category) but are not acquired by age 2;6. Except for three adjectives (*naughty, poor, last*) and one people word (*child*), all are closed-class words, including pronouns, time words, quantifiers, articles, prepositions and question words. The most frequent words in this group were *was* (11,495 occurrences) and *of* (16,529). These results are consistent with the finding discussed earlier that the high frequency of closed-class words does not lead to early acquisition. As noted earlier, another potentially troublesome category, that of words that are acquired by age 2;6, but are not present in the input data, does not occur at all.

Summing up the production findings, we see that contrary to prediction, higher frequency in parental input is associated with later age of acquisition of specific words when the entire lexicon is examined. However, within lexical categories, higher frequency is related to earlier age of acquisition as hypothesized. Furthermore, this relation is much stronger when based on the frequency of words in oral, child-directed speech rather than written language directed toward adults or older children.

Comprehension data is available for children aged 0;8 to 1;4 based on a smaller set (157) of words. The relevant correlations with parental

TABLE 3. *Correlations between age of acquisition in comprehension and selected estimates of word frequency*

Category and number of words	*r* (parental frequency)	*r* (Kucera–Francis)	*r* (Thorndike–Lorge)
Common nouns (79)	0·20*	0·08	0·04
People words (6)	0·06	−0·64	−0·62
Verbs (34)	−0·14	−0·14	0·05
Adjectives (10)	0·17	0·56	0·57
Closed class (10)	0·04	−0·26	−0·30
Others (18)	0·37	0·33	0·50*

* $p < 0{\cdot}05$, ** $p < 0{\cdot}01$.

frequency, along with the number of words in each category, are shown in Table 3. The correlations are generally smaller than those for production, and only the relation between age of acquisition and parental frequency for common nouns is significant. One possible reason for the lack of a significant relation between age of acquisition and parental frequency for the other five lexical categories is that the number of words in each category is smaller, due to the shorter list on the CDI:WG, which is focused on an earlier stage of development. For example, the apparently substantial correlations for adjective age of acquisition with Kucera–Francis and Thorndike–Lorge norms are based on just ten words. We also note that the interpretation of significance testing of correlations is unclear in the present context. Significance testing is most relevant for estimating the generalizability of the results from a sample to a larger population. In the present case, the vocabulary items available, while not entirely exhaustive, do constitute the largest proportion of words regularly acquired in this age range. The correlations themselves are more usefully viewed as descriptive statistics that estimate the effect size of frequency.

A final question is whether the effects of frequency change with development, either increasing or decreasing as more words are acquired. To evaluate this possibility, we returned to the production data and divided the words into those typically acquired before and after children produce 100 words. In the norming study from which we draw this data, the mean age at which children produced 100 words was 19·2 months. We chose 100 words because the rate of vocabulary acquisition commonly increases when children know between 50 and 100 words. Thus, by the time children know 100 words, they are learning words more quickly, and the effect of any one variable, such as frequency, may have changed from the initial phases of vocabulary acquisition. We then assessed whether parental frequency and age of acquisition (in production) were related differently for the early

TABLE 4. *Correlations between age of acquisition in production and selected estimates of word frequency for words acquired early (in first 100) or later (after first 100)*

Category	Words acquired in first 100: *r* (*N*)	Words acquired after first 100: *r* (*N*)
Common nouns	0·19 (53)	0·43** (203)
People words	0·85* (5)	0·60** (16)
Verbs	0·26 (4)	0·14 (86)
Adjectives	0·36 (3)	0·27* (52)
Closed class	0·82 (4)	0·38** (64)
Others	0·13 (10)	0·14 (62)

* $p < 0·05$, ** $p < 0·01$.

period of acquisition (first 100 words) and for the later period of acquisition (post-100 words). As shown in Table 4, the correlations show that there is a substantial change in correlation for two categories, nouns and closed-class words. For nouns, the correlation with frequency was higher for words after the first 100. That is, frequency has little effect within the first 100 words for nouns, but becomes more important later. In contrast, frequency appears to be more important (though non-significant) for very early closed-class words and somewhat less important later. However, this aspect of the results, particularly the high correlation for the early words, is probably due to the very small number of words acquired prior to age 1;7: *down* and *up* are highly frequent (6,899 and 11,477, respectively) and acquired prior to age 1;5, whereas *mine* and *more* are somewhat less frequent (669 and 4,197, respectively) and acquired after age 1;6.

DISCUSSION

In the present study, we have provided the first comprehensive evaluation of the role of parental input frequency in early (through 30 months) vocabulary acquisition. Three aspects of the results are notable. The first is the role of specific semantic–syntactic categories. The correlation between input frequency and age of acquisition for the 562 words as a whole is contrary to that hypothesized. Nouns are least frequent individually, but learned the earliest. Indeed, a noun bias in early language development is widespread if not quite universal. Thus, the present results show that factors other than frequency are responsible for the ease of acquisition of nouns, at least in English. Identifying those factors remains an important topic of debate (Gentner, 1982; Tardif *et al.*, 1999). Closed-class words are the most frequent items individually, yet the slowest to be acquired.

It is important to note that although we have grouped the words into lexical categories used by the CDI as well as other studies, we are not claiming that children themselves classify the words, tacitly or otherwise, as nouns, verbs, adjectives and so on. Further, following the CDI, we have classified words that adults would treat as nouns in more than one category (common nouns, people words, other). The categories are simply meant to represent 'bundles' of syntactic, semantic and phonological features in the input which children may utilize in their learning. Indeed, many factors will ultimately affect age of acquisition. For example, the role of semantic concreteness, syntactic complexity, informational load and ease of perception may all contribute to the striking fact that overall, the most frequent words within this particular set of 562 words are learned later. The important point here is that the effect of frequency on vocabulary acquisition interacts with semantic–syntactic categories.

Although across the vocabulary as a whole, highly frequent words are not learned earlier, WITHIN each of the six categories, the more frequently the word is heard, the earlier it is acquired in expressive vocabulary. Restriction of the correlations to the words in individual categories has the effect of increasing homogeneity on other variables that may influence age of acquisition. These variables may include perceptibility of the referent of the word, diversity of syntactic frames, requirements for interpersonal understanding, use in joint attention contexts and others. Each vocabulary category might be thought of as defining its own 'problem space', and within that space, frequency appears to play a significant role. However, the strength of the relationship between frequency and age of acquisition varies considerably across categories. The strongest relation is found for nouns. The nouns learned in this early stage of development are primarily concrete and basic, and as a result, their semantic–syntactic diversity may be the smallest of the categories, permitting frequency to play a large role. In contrast, the weakest relations are found for verbs and closed-class words, which may be more variable categories. The closed-class category aggregates several categories, such as particles, auxiliaries, determiners and prepositions, and therefore has a great deal of semantic and syntactic diversity; Bornstein *et al.* (2004) make a similar argument for verbs. Overall, although frequency plays a substantial role, there is much variance left to be explained by other variables such as those listed above. We suggest that such research should always begin by first partialing out the role of frequency within each category. The residual may be conceptualized as a measure of the difficulty of the word, providing a basis for examining the role of phonological, grammatical and semantic properties as determinants of word difficulty.

A second notable aspect of the results is the demonstration that age of acquisition is more strongly related to parental frequency as estimated

from transcripts of parent–child interaction than from norms based on adult–adult communication or written materials prepared for older children. Although this is hardly a surprising result, the magnitude of the effect is substantial for production. This finding is significant because tests of frequency effects on language acquisition and language processing in young children rarely use frequency norms designed for that age group. The results might be quite different if appropriate frequency norms were used. Furthermore, the present study provides confirmation that estimates of parental frequency from the CHILDES transcripts are valid estimates of frequency. The total sample they are based on (3·8 million words) is somewhat larger than the corpus for adult-directed written language (just over one million words for the Kucera–Francis norms) and smaller than the corpus for written language for older children (over 13·5 million words for the Thorndike–Lorge norms).

Finally, the present findings suggest a difference between comprehension and production with respect to the role of input frequency. Parental frequency is a substantially more consistent predictor of age of acquisition for production than for comprehension. This difference might reflect lower validity for parental reports of vocabulary comprehension on the CDI. However, a recent review of CDI validity research (Fenson *et al.*, 2007) suggests that comprehension scores are only slightly less valid than production scores. Another possible explanation is that the difference is due to the inclusion of later acquired words in the production data relative to the comprehension data, together with a developmental shift toward a greater role for frequency. However, with the exception of common nouns, little evidence was found for such a shift within the production data (cf. Table 4). Thus the stronger predictive effect of parental frequency for production than for comprehension does not appear to be an artifact of stage of development. We suggest a probabilistic explanation of this somewhat counter-intuitive finding. As a first approximation, assume a small but fixed number of exposures to a word are sufficient for an initial level of comprehension, whereas a larger but also fixed number of exposures are required to form a sufficiently detailed representation for production. Also assume that instances of a specific word are randomly distributed in the input. It then follows from stochastic probability models that the time it takes to accumulate the larger number of exposures will be more highly correlated with input frequency than the time it takes to accumulate the smaller number. Consequently, processes that occur later in time, such as production, will show a stronger relationship to frequency than those that occur earlier, e.g. comprehension.

The results just discussed concerning comprehension and production also bear on a particularly challenging question, that of inferring direction of causality. Most researchers have argued, or at least assumed, that frequency

is a positive, causal facilitator of lexical acquisition. Lexical training studies can be adduced in support of this interpretation. But like all correlational results, the present findings may reflect a variety of underlying mechanisms. One plausible alternative reverses the direction. Perhaps parents use certain words more because they believe their children understand them better than other words. This interpretation would predict a higher correlation with frequency for comprehension than for production, which is not the case. A variant of this hypothesis is that the appearance of a word in the child's production stimulates higher use by the parent. Thus at any given point in development, those words which the child is producing will be produced with higher frequency by the parent. In order to evaluate this hypothesis a fairly precise identification of the first appearance of the word is needed, and therefore it is only possible with highly 'dense' datasets of parent–child interaction. Finally, it is possible that 'third factors' are causally responsible for both input frequency and child acquisition. The most likely candidates are features of linguistic complexity, whether phonological, syntactic, semantic or pragmatic, which might independently make parents less likely to use them and also delay acquisition. However, although these features might be responsible for the positive correlations found within the lexical categories, they cannot explain the negative correlations found across categories. It is possible that more complex training studies might be devised to distinguish these possibilities, which are, of course, not mutually exclusive.

In summary, the answer to the question 'Does frequency count?' is 'Yes', but the way it counts is not straightforward. It clearly depends on the type of words being acquired (e.g. nouns vs. other lexical categories), the modality of acquisition (production vs. comprehension), and the time line of acquisition (earlier vs. later role of frequency). In addition, frequency is clearly only part of the story. Thus, while it is an important piece of the puzzle of which words children will learn when, the way it interacts with other variables needs to be further explored.

REFERENCES

Barrett, M., Harris, M. & Chasin, J. (1991). Early lexical development and maternal speech: a comparison of children's initial and subsequent uses of words. *Journal of Child Language* **18**, 21–40.

Bates, E., Dale, P. S. & Thal, D. (1995). Individual differences and their implications for theories of language development. In Paul Fletcher & Brian MacWhinney (eds), *Handbook of child language*, 96–151. Oxford: Basil Blackwell.

Benedict, H. (1979). Early lexical development: comprehension and production. *Journal of Child Language* **6**, 183–200.

Blackwell, A. A. (2005). Acquiring the English adjective lexicon: relationships with input properties and adjectival semantic typology. *Journal of Child Language* **32**, 535–62.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S., Pascual, L., Pecheux, M., Ruel, J., Venuti, P. & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young

children: Spanish, Dutch, French, Hebrew, Italian, Korean and American English. *Child Development* **75**, 1115–39.

Brent, M. R. & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition* **81**, B33–B44.

Brown, R. (1973). A first language: the early stages. Cambridge, MA: Harvard University Press.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller & M. Halle (eds), *Linguistic theory and psychological reality*, 264–93. Cambridge, MA: MIT Press.

Choi, S. & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language* **22**, 497–529.

Clark, E. V. & Hecht, B. F. (1983). Comprehension, production and language acquisition. *Annual Review of Psychology* **34**, 325–49.

Dale, P. S. & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers* **28**, 125–27.

Dromi, E. (1987). *Early lexical development*. Cambridge: Cambridge University Press.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59** (5, Serial No. 242).

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S. & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: user's guide and technical manual*, 2nd ed. Baltimore: Paul H. Brookes.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. A. Kuczaj (ed.), *Language development: Vol. 2. Language, thought and culture*, 301–334. Hillsdale, NJ: Erlbaum.

Goldfield, B. A. & Reznick, J. S. (1990). Early lexical acquisition: rate, content and the vocabulary spurt. *Journal of Child Language* **17**, 171–83.

Golinkoff, R. M., Hirsh-Pasek, K., Bloom, L., Smith, L. B., Woodward, A. L., Akhtar, N., Tomasello, M. & Hollich, G. (2000). *Becoming a word learner: a debate on lexical acquisition*. New York: Oxford University Press.

Gopnik, A. & Choi, S. (1990). Do linguistic differences lead to cognitive differences? A cross-linguistic study of semantic and cognitive development. *First Language* **10**, 199–215.

Harris, M., Barrett, M., Jones, D. & Brookes, S. (1988). Linguistic input and early word meaning. *Journal of Child Language* **15**, 77–94.

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: relation to language input and gender. *Developmental Psychology* **27**, 236–48.

Kucera, H. & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Li, P., Burgess, C. & Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. In E. V. Clark (ed.), *Proceedings of the Thirtieth Annual Child Language Research Forum*, 167–78. Stanford, CA: Center for the Study of Language and Information.

Li, P., Zhao, X. & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science* **31**, 581–612.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk*, 3rd ed. Mahwah, NJ: Erlbaum.

Naigles, L. R. and Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* **25**, 95–120.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monograph of the Society for Research in Child Development* **38**, 1–2, Serial #149.

Patterson, J. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics* **23**, 493–508.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J. & Pae, H. K. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech & Hearing Research* **37**, 106–121.

Schwartz, R. G. & Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of Child Language* **10**, 57–64.

Tardif, T., Gelman, S. A. & Xu, F. (1999). Putting the 'noun bias' in context: a comparison of English and Mandarin. *Child Development* **70**, 620–35.

Tardif, T., Shatz, M. & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian and Mandarin. *Journal of Child Language* **24**, 535–65.

Thorndike, E. L. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Wang, M. & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning* **55**, 71–98.

Weizman, Z. O. & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology* **37**, 265–79.