
Volume 63, Issue 3, October 2010
ISSN 0749-596X
ELSEVIER
Journal of
Memory and
Language
Editor
Victor Ferreira
Associate Editors
Sally Andrews
Simon Farrell
Robert Greene
James Magnuson
Editorial Board
Available online at www.sciencedirect.com
ScienceDirect

Journal of Memory and Language 63 (2010) 259–273

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The associative structure of language: Contextual diversity in early word learning

Thomas T. Hills [a,*], Josita Maouene [b], Brian Riordan [c], Linda B. Smith [d]

[a] Cognitive and Decision Sciences, University of Basel, Switzerland
[b] Department of Psychology, Grand Valley State University, United States
[c] Aptima, Inc., United States
[d] Department of Psychological and Brain Sciences, Indiana University, United States

ARTICLE INFO

Article history:
Received 5 August 2009
revision received 4 June 2010

Keywords:
Network analysis
Age of acquisition
Child-directed speech
Free association
Contextual diversity
Word learning

ABSTRACT

Previous studies demonstrated that statistical properties of adult generated free associates predict the order of early noun learning. We investigate an explanation for this phenomenon that we call the associative structure of language: early word learning may be driven in part by contextual diversity in the learning environment, with contextual diversity in caregiver speech correlating with the cue–target structure in adult free association norms. To test this, we examined the co-occurrence of words in caregiver speech from the CHILDES database and found that a word's contextual diversity—the number of unique word types a word co-occurs with in caregiver speech—predicted the order of early word learning and was highly correlated with the number of unique associative cues for a given target word in adult free association norms. The associative structure of language was further supported by an analysis of the longitudinal development of early semantic networks (from 16 to 30 months) using contextual co-occurrence. This analysis supported two growth processes: The lure of the associates, in which the earliest learned words have more connections with known words, and preferential acquisition, in which the earliest learned words are the most contextually diverse in the learning environment. We further discuss the impact of word class (nouns, verbs, etc.) on these results.

Introduction

Why do children learn some words earlier than others? Developmentalists have offered answers in terms of the psychology of young children pointing to their interests (e.g., Goldin-Meadow, Goodrich, Sauer, & Iverson, 2007) and the early cognitive availability of some concepts over others (e.g., Markman, 1989; Waxman & Markow, 1995); but researchers are also increasingly offering answers in terms of the statistical properties of the language-learning environment, pointing to the frequency of individual words in the input (Goodman, Dale, & Li, 2008), the diversity of the contexts in which they appear (e.g., Hurtado, Marchman, & Fernald, 2008), and the density and sparseness of the semantic connections between words in the input (Hills, Maouene, Maouene, Sheya, & Smith, 2009a). Why, in language processing, do adults more rapidly access some words over others? Researchers have documented a variety of relevant statistical factors including the frequency of individual words in the language (e.g., Murray & Forster, 2004), the diversity of the contexts in which they appear (Adelman, Brown, & Quesada, 2006), the density and sparseness of semantic connections to other words (Griffiths, Steyvers, & Firl, 2007), and the age of acquisition of the words (Gilhooly & Gilhooly, 1979). All this suggests potentially meaningful correspondences between the statistical properties that matter for early lexical learning and those that matter later for efficient lexical processing.

* Corresponding author. Address: University of Basel, Missionsstrasse 64a, 4055 Basel, Switzerland. Fax: +41 61 267 04 41.
E-mail addresses: thomas.hills@unibas.ch, thomhills@gmail.com (T.T. Hills).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2010.06.002

This paper examines the relation between one index of the semantic connectivity of words in people's heads and the statistical properties of the external language-learning environment that may mediate this relationship. The index of semantic relatedness is associative strength and it has been shown repeatedly to be predictive of lexical processing and retrieval in adults (e.g., Nelson, Schreiber, & McEvoy, 1992; Steyvers, Shiffirin, & Nelson, 2005) and more recently, to also be predictive of the order of acquisition of nouns by young children (Hills et al., 2009a). Associative strength is itself an index derived from studies in which researchers provide word cues (e.g., dog) and participants give back the first word that comes to mind (e.g., cat). Associative strength is thus a property of large data sets of answers given back by the individual participants in an odd laboratory task. Presumably these association norms reflect in some aggregate way important regularities in language, language use, and the world (Deese, 1965). It would seem that they must in that adult association norms are strongly predictive of adult performance in many different kinds of language processing tasks (e.g., Nelson et al., 1992). Still, adult association norms are best considered a statistical portmanteau—holistically useful but without a well-articulated link to the psychological mechanisms or the substructure of the language they carry. They may for example represent semantic relationships—the shared semantic context of "soap" and "bath"—and they may just as easily reflect compound words, where semantic relationships are less precise—like "bird" and "bath." This makes the fact that adult association norms also predict the order of acquisition of words by toddlers intriguing and potentially revealing.

This paper provides an analysis of the contextual diversity of relations among words in the language-learning environment of young children (i.e., their co-occurrence in child-directed speech) in relation to adult association norms and in relation to the normative age of acquisition of those words. The rationale for the analyses presented in this paper derives from our past work focusing on how the structure of semantic networks may characterize developing vocabularies. Our hypotheses also follow from the evidence concerning lexical diversity as a relevant statistical property in acquisition and lexical processing, and from recent hypotheses about the possible growth patterns of semantic networks (density of connectivity of the words that have already been learned versus that of the words not yet acquired). In what follows, we first describe this rationale in more detail before moving on to describe our specific approach.

Early semantic networks

Graph theory, or network analysis, can be applied to any structure that consists of nodes connected to each other through links or edges. For example, nodes might be cities and links might be roads; or nodes might be proteins and links might be the molecules that bind with and activate them; or, nodes might be words and the links indices of semantic connectedness such as cue–target associations in free association norms or frequency of co-occurrence in language corpora. With the advance of graph theory, there have been increasing analyses of the structure of language in terms of semantic networks (e.g., Griffiths et al., 2007; Steyvers & Tenenbaum, 2005). The semantic networks used in these analyses may be built from various sources, including corpora collected from written or spoken language, free association data, and feature-norms that indicate shared perceptual and conceptual properties (e.g., Hills et al., 2009a; Hills, Maouene, Maouene, Sheya, & Smith, 2009b; Steyvers & Tenenbaum, 2005). Semantic networks built in these ways from adult normative data have structural properties that are believed to support efficient processing and word retrieval (Griffiths et al., 2007; Steyvers & Tenenbaum, 2005). One property of large scale semantic networks that has been linked to their growth processes is a power-law structure in the distribution of links across nodes: most nodes in the network have few links to other nodes but a few nodes are hubs, having many links to other nodes. The power-law distribution of links over nodes has been shown to emerge when networks grow following the principle of preferential attachment (Barabasi & Albert, 1999; Steyvers & Tenenbaum, 2005). By this principle, nodes with many links (what are called high-degree nodes) are more likely to add new links than nodes with fewer links.

In two previous studies (Hills et al., 2009a,b), we examined the developing structure of children's semantic networks in an attempt to understand how the connectivity among words might be related to lexical growth. The networks were built from the normative vocabularies of young children learning English at different ages. These normative vocabularies were taken from the Bates-McArthur Communicative Developmental Inventory (CDI), a large study of the vocabularies of young children, which is commonly used to assess levels of vocabulary development. Critical to our purposes, these month-by-month acquisition norms specify the words that are in the productive vocabulary of 50% of children at a given month of age. Thus, we can access the set of words that are normatively known by different aged children learning English. These words are the nodes in the semantic networks. But how does one link them to build a network that reflects semantic connectivity?

In our previous work, we used only early-learned nouns from the CDI and connected them to each other to build networks either by the association strength of the nouns in *adult* association norms or by shared perceptual and functional features from *adult*-generated feature norms (McRae, Cree, Seidenberg, & McNorgan, 2005). Although the adult-generated features captured relevant aspects of children's developing knowledge of superordinate categories, the networks generated from shared features did not show a power-law structure and connectivity in these networks was not strongly related to age of acquisition (Hills et al., 2009a). In contrast, adult association norms did yield a power-law structure and the number of links received by a word from other words (indegree) was strongly related to age of acquisition (Hills et al., 2009a). Children, of course, do not have direct access to information about adult free associates and so the fact that adult association norms predict age of acquisition is provocative. Adult free associations—as a portmanteau index of semantic relatedness—must be capturing some relevant aspect of the learning environment for young children.

Contextual diversity

One statistical property of associations that may be related to an important property of the lexical learning environment is the diversity of other words that elicit the target noun as an associate. For example, from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998), "ball", "shoe", and "toy" are produced as a target of 83, 28, and 17 unique cue words, respectively. At 16 months, according to the CDI (Dale & Fenson, 1996), the same words are, respectively, produced by 85%, 59%, and 27% of the children studied. In general, words recalled as targets for a larger set of cues are acquired earlier in development than words recalled for a smaller set of cues; this is true even after controlling for frequency and aspects of phonology (Hills et al., 2009a; Steyvers & Tenenbaum, 2005).

This suggests the specific hypothesis examined in the present study, which we call the *associative structure of language*: Adult association norms are correlated with the order of word acquisition by children because adult association norms are correlated with the co-occurrence statistics of words in the learning environment. Contextually diverse words, which co-occur with many other word types, are acquired earlier in development than less diverse words, and in adults these contextually diverse words are also offered as the associates (the targets) of many other words in free association tasks.

Diversity of linguistic contexts is itself associated with better speech perception and segmentation in both infants and adults (e.g., Hayes & Clark, 1970; Newman, 2008; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). Contextual diversity is also correlated with word naming and lexical decision times in adults (see Adelman & Brown, 2008; Adelman et al., 2006; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Steyvers & Malmberg, 2003; Verkoeijen, Rikers, & Schmidt, 2004). Still other studies show that more varied and complex input benefits performance in adult artificial-language learning tasks (Gillette, Gleitman, Gleitman, & Lederer, 1999; Plaut & Kello, 1999: Recchia, Johns, & Jones, 2008; Yu & Smith, 2007). Finally, some studies suggest that increased lexical diversity in parent speech leads to a more diverse early vocabulary, more rapid vocabulary growth, and better word recognition by young children (e.g., Hoff & Naigles, 2002; Hurtado et al., 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Naigles & Hoff-Ginsberg, 1998; Rowe, 2008). Many of these observations are consistent with the Syntactic Bootstrapping Hypothesis, in which the words that surround a word in language provide information about the meaning of that word (Gleitman, 1990). However, none of the above studies show that the contextual diversity of individual words in the input predicts their age of acquisition or that this is connected with an associative structure in the language—the questions we address here.

Importantly, some analyses of language acquisition have suggested the opposite relation between contextual diversity and acquisition: that consistency of context rather than diversity benefits early learning (Goldberg, Casenhiser & Brown, 2007; Meints, Plunkett, & Harris, 2008; Mervis, 1987; Sethuraman, 2004). For the most part, these hypotheses concern verb learning (and possibly adjective learning, Mervis, Mervis, Johnson, & Bertrand, 1992; Waxman & Klibanoff, 2000), and not nouns, the lexical domain in which prior work indicated that the number of adult associates predicted the age of acquisition of early-learned nouns (Hills et al., 2009a). One specific result suggesting a beneficial effect of context consistency is the strongly conservative character of children's early verb use; individual verbs are produced only in a narrow range of syntactic and possibly also semantic contexts (e.g., Akhtar & Tomasello, 1997; Goldberg, Casenhiser, & Sethuraman, 2004). Other evidence suggests that the syntactic and semantic contexts of verb use in caregiver speech are also restricted—favoring consistency over diversity at least by some metrics (Goldberg et al., 2004; Maouene, Laakso, & Smith, submitted for publication; Sethuraman, 2004). There is also evidence that early consistency in the syntactic contexts of verbs may promote more rapid learning (Goldberg et al., 2004; Sethuraman, 2004; see also Meints et al., 2008; but see Hoff & Naigles, 2002; Naigles & Hoff-Ginsberg, 1998). In the domain of adjective learning, Mervis et al. (1992) reported that children learned color words more rapidly when they were first used in highly consistent contexts. Sandhofer and colleagues (Sandhofer & Doumas, 2008; Sandhofer & Smith, 2007) also found that teaching color words was more successful using a progress-alignment approach in which contextual consistency was followed by contextual diversity. Finally, in a series of programmatic experiments, Waxman and Klibanoff (2000) showed that adjectives are better learned by young children in contexts in which they modify the same basic-level noun category rather than ones in which they modify different nouns.

Thus, we have on the one hand, a reasonable hypothesis about how and why adult free association norms correlate with age of acquisition of *nouns* by children. The hypothesis is that these adult free association norms also correlate with the contextual diversity of words in the learning environment, such that words that are frequent associates of other words also co-occur with many other words in the learning environment (and presumably also with extra-linguistic contexts). This contextual diversity is hypothesized to help children learn these words and their meanings. However, on the other hand, some developmental evidence seems to suggest that for verbs and adjectives, consistency rather than diversity is related to early acquisition. In the present study, we will examine the co-occurrence structure of nouns, verbs, adjectives, and function words in speech to young children in relation to the age of acquisition of those words and in relation to adult free association norms.

Growth patterns in semantic networks

Within a network representation based on word co-occurrence, words that co-occur with many other words will be high-degree nodes, and potential hubs in the network. Thus, contextual diversity may be related to network growth. More specifically, the power-law distribution of links over nodes may emerge as a consequence of *preferential attachment* (Barabasi & Albert, 1999; Steyvers & Tenenbaum, 2005), whereby nodes with many links are more likely to add new links than low degree nodes. This implies that words with many links, contextually diverse words, will not only be early but that words that link to these high-degree words will also be earlier than words that do not. Steyvers and Tenenbaum (2005) proposed that early word learning followed this principle of preferential attachment, with networks growing by adding new words based on their connectivity to well connected words already in the growing network.

Our previous analyses of noun vocabulary growth suggested that there were alternate growth principles that might be more accurate for early word learning (Hills et al., 2009a). One alternative principle is what we termed *preferential acquisition*, in which more well connected nouns in the learning environment (rather than more connected nouns in the child's internal network as in preferential attachment) are learned earlier than less well connected nouns (Hills et al., 2009a). Yet another alternative principle is called the *lure of the associates* (Hills et al., 2009a). With the lure of the associates, new words are added to the lexicon in direct proportion to the number of known words that they are related to (e.g., knowledge of words about animals invites in new animal related words). Where preferential attachment is based on the connectivity of known words with other known words, preferential acquisition is the based on the connectivity of words in the learning environment overall, and the lure of the associates is based on how many known words are related to newly learned words (see Fig. 1).

Our prior study of network growth found the strongest support for preferential acquisition, but it considered only 130 early-learned nouns and built developing networks using adult free association data. Here, we consider a bigger set of early-learned nouns, as well as non-nouns, and construct developing networks using the common contexts of words in child-directed speech. Discriminating among these different possible growth patterns should provide insight to why contextual diversity, as a property of an emerging semantic network rather than a property of an individual word, might support learning.

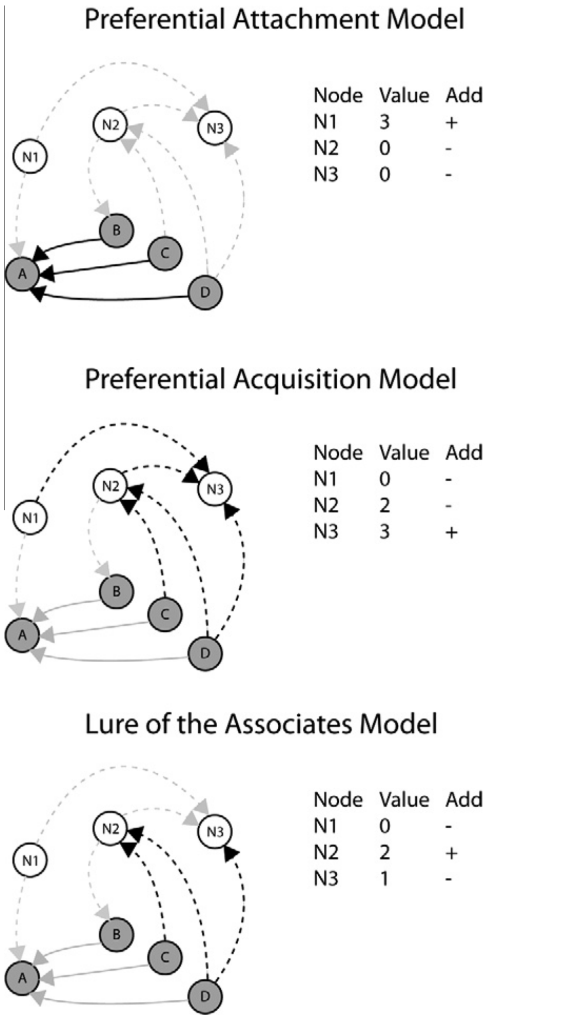

Fig. 1. The three growth models presented for a simplified network. Each of the networks is the same, but the growth models assign different learning values to each unknown node (i.e., word). Possible new nodes are shown in white, known nodes are represented in gray; links relevant to the growth model are shown in black, unimportant links in gray. "Add" indicates which of the three unknown nodes is favored for learning by the growth model. Learning values provide a quantitative measure of the learning (or growth) potential for a given node. With preferential attachment, the value of the new node is the average degree of the known nodes it would attach to. With preferential acquisition, the value of the new node is its degree in the full network (which includes what is known and not known). With the lure of the associates, the value of new node is its degree with respect to known nodes. Arrows indicate a directed network, but the same models apply for undirected networks. This figure is taken from Hills et al. (2009a).

The present study

The primary goals of this study are threefold. The first goal seeks to understand the relation between adult free associates for nouns and their age of acquisition. If the number of associates for a noun is an index of the diversity of linguistic contexts in the input, and it is this diversity that is responsible for the relation between adult associates and age of acquisition, then the age at which a noun is acquired should be a consequence of its contextual diversity in child-directed speech. In this paper, we use a corpus of child-directed speech from the CHILDES database (MacWhinney, 2000) to investigate whether the contextual diversity of words in child-directed speech correlates with their age of acquisition. In addition, we ask if contextual diversity might be a better predictor of age of acquisition than the number of cues a given target noun has in the free association norms—with the implication that the ability of free association norms to predict age of acquisition is explained by their relationship to contextual diversity in child-directed speech. The second goal is to examine non-noun word classes. Different kinds of words present different learning tasks and may be learned in fundamentally different ways (e.g., Gentner, 1978; Goodman et al., 2008; Sandhofer, Smith, & Luo, 2000). Thus, contextual diversity might help learning some kinds of words but not others. The third goal is to examine contextual diversity in relation to different learning processes—such as preferential acquisition, preferential attachment, and the lure of the associates.

Study 1: regression analyses of age of acquisition using free associates and contextual diversity in child-directed speech

These analyses are specifically directed to whether (1) the predictive relation between adult free associates and the normative age of acquisition of early nouns might be understood in terms of a relation between contextual diversity in the learning environment and age of acquisition and (2) whether the correlation between free associates and age of acquisition and the expected correlation between contextual diversity in the input and age of acquisition hold for other kinds of words than nouns.

Method

Words

The words are those on the CDI (Dale & Fenson, 1996), Toddler version. This inventory is a checklist of 680 words and phrases typically acquired at a young age by children learning English as a native language. The CDI includes data on the normative productive vocabularies of children in one-month increments from 16 to 30 months of age. For example, at 16 months, an average of 25.6% of children produce the word "pig" in their speech. By 30 months, 92.9% use the word. For our analyses, we excluded words about time (representing a class of 12 words with occurrences too small for reliable analyses), games and routines (which were generally not single words), sound effects and animal sounds, words that were duplicated across categories (e.g., "orange" is in the category of foods and colors), and 42 words that were never recalled as target words in the free association norms. This left 532 words, of which 330 were nouns, 96 were action words (verbs), 58 were descriptive words (adjectives), and 88 were function words consisting of pronouns, quantifiers, articles, helping verbs, and connecting words. Age of acquisition for a word was defined as the first month at which the word was produced by more than 50% of the children in the normative sample.

Associates

We used the adult-generated University of South Florida Free Association Norms (Nelson et al., 1998). These were collected by providing adult participants with a word (the cue) and asking them to provide the first word that came to mind (the target). This establishes a cue–target relationship. For example, when provided with the cue word "cat" many participants provide the target word "dog". The norms consist of approximately 5000 cue words and their related targets. The number of associative relationships was taken as the count of the number of distinct cue words for which the target word was recalled. For clarity, we will call this value the *associative indegree*.

CHILDES

The co-occurrences of words in the language-learning environment were assessed through analyses of a 2 million word corpus of caregiver speech derived from the CHILDES database (MacWhinney, 2000). The exact corpus is the same as that used by Riordan and Jones (2007) and consists of caregiver speech from the American section of the CHILDES database. The speech derives from many different adult-child interactions including toy play, reading, and conversation. The speech in this sample was directed to children from 12 to 60 months. These samples have been contributed over the years by many different researchers and thus there is some variation in transcription conventions. Our analysis used Riordan and Jones (2007) standardization of the transcriptions. Analyses of the corpus were conducted, as described below, for the 532 words from the CDI. Variants of the same stem were treated as tokens of the same CDI word (e.g., cats → cat) as they are in the normative standards for acquisition on the CDI.

Contextual diversity

Our measure of contextual diversity starts by building a matrix of word co-occurrences using a process similar to the Hyperspace Analogue to Language (HAL) (Lund & Burgess, 1996) and the word co-occurrence detector (Li, Farkas, & MacWhinney, 2004). For a corpus of N unique word types, an $N \times N$ matrix is formed, where each cell, ij, is filled according to the following rule: a moving window of size k moves word-wise through the corpus, with the *initial* word, i, adding a unit of 1 to cell ij, if the word j is in the window simultaneously with i. Table 1 presents an example for the sentence "The ball is on the dog".

Using this method, we generated separate matrices for a series of window sizes ranging from 2 to 100. Each matrix was then converted to a binary matrix (i.e., nonzero entries were truncated to 1). Therefore, each cell ij represented the co-occurrence of word i with word j within a window of k words throughout the corpus. Prior to converting to a binary matrix, the sum of the cells for a given word is strongly correlated with its frequency.

Summing across rows and down columns gives the number of different unique word contexts for a given window size, and thus provides a straightforward indication of a word's contextual diversity. In network terms, this sum is equivalent to the *degree* of a word—measuring both the number of unique word types that come immediately before and the number of unique word types that follow a given word. In the following analyses we call this measure of diversity the *CHILDES degree* (or CHd). This operationalization of contextual diversity differs from previous studies of contextual diversity, lexical decision, and

Table 1
Sample co-occurrence matrix for the sentence "The ball is on the dog" using a window of size 3. For example, for the word "the", the words "ball", "is", and "dog" follow within two words, and are therefore in a three-word window with "the". "Ball" appears in a three-word window with "is" and "on", and so forth.

	the	ball	is	on	dog
the	0	1	1	0	1
ball	0	0	1	1	0
is	1	0	0	1	0
on	1	0	0	0	1
dog	0	0	0	0	0

word naming based on diversity across documents (Adelman et al., 2006; Recchia et al., 2008). The present method may be more appropriate for these analyses because conversations in the CHILDES corpus are of varying size and because it is unclear *a priori* what the proper resolution of the attentional frame should be for a young listener. The analyses reported below suggest that the best window size for detecting contextual diversity predictive of future word learning is between 5 and 50 words, much smaller than the typical document size in other measures of contextual diversity. Frequency counts were taken as the number of occurrences of a given word in the corpus.

Statistical analyses

Statistical analyses were performed in R (R Development Core Team., 2009). The p-values for the regressions computed in Tables 2–4 are based on the ANOVA likelihood ratio test. When controlling for another factor (e.g., log-AI after log-WF), the p-value and R^2 is based on the incremental improvement in the error sum of squares as the new factor is added to a regression model that already contains the control factor.

Results and discussion

To investigate the relationship between associative indegree and age of acquisition for different word classes, we performed a series of linear regressions with AoA as the dependent variable using the full set of 532 words. As the first column of Table 2 shows, associative indegree is predictive of AoA across all word classes. Fig. 2 displays the standardized regression coefficients for the same regressions, indicating that in all cases, words that are learned earlier have a significantly higher associative indegree than words that are learned later. Next, we compared the variance explained by associative indegree with that explained by word frequency. Word frequency is also predictive of AoA (Table 2, second column). However, for nouns, verbs, and function words, associative indegree explains variance above and beyond that which is explained by word frequency alone (Table 2, third column).

These findings confirm and extend our previous findings (Hills et al., 2009a), demonstrating that associative indegree is a significant predictor of age of acquisition for a broader class of nouns, as well as other word classes, during the first 30 months of word learning. In all cases, the sign of the coefficient for log-AI was negative, indicating that words that are the target for a larger set of cues (have more free associates) are more likely to be learned earlier than words associated with a smaller set of cues.

Might associative indegree predict age of acquisition because it is also associated with contextual diversity in the learning environment? One way to approach this hypothesis is to ask what aspects of the structure of child-directed language are similar to adult generated free associates, and whether these aspects are also predictive of age of acquisition. To investigate the correlation between contextual diversity of words in child-directed speech and AoA, we used the co-occurrence statistics of caregiver speech to compute the CHILDES degree (CHd). The CHd measures the diversity of semantic contexts that a particular word appears in. We then used the CHd as an independent variable, and controlling for frequency of the word in the CHILDES corpus, we regressed these on the AoA from the CDI for each word in a given word class. For these analyses, we used multiple window sizes ranging from 2 words to 100 words, and from the resulting statistics computed the CHd.

Fig. 3 presents the results of the regression analyses, for window sizes ranging from 2 to 100. For word classes excluding nouns, contextual diversity computed in terms of smaller window sizes is more predictive of age of acquisition, and this peaks at a window size around 5. This means that for most words, the contextual diversity relevant to learning is within a relatively short span of about 5 words. However, the optimal neighborhood for computing contextual diversity (so as to predict age of acquisition) may vary with syntactic category: nouns, and to some extent adjectives, benefit from co-occurrence statistics in larger windows and over multiple sentences in caregiver speech. This is intriguing and, as we discuss in the general discussion, potentially relevant for how children learn different kinds of words. Also, as shown below, without controlling for frequency, nouns have the highest correlation between age of acquisition and contextual diversity. Because a window of 5 is a reasonably good predictor for nouns and the best for other words, we use a window size of 5 in all further analyses.

Table 2
Effects of log-transformed associative indegree (log-AI) and log-transformed word frequency (log-WF) on age of acquisition for nouns, verbs (action words), adjectives, and function words (e.g., determiners, prepositions). ΔR^2 represents the change in the R^2 of the linear regression that is a consequence of the relevant measure. Log-AI = the log-transformed associative indegree. Log-WF = log-transformed word frequency. "After" indicates the additional change in R^2 correlated with the first measure after controlling for the second. In all cases the regression coefficients are negative, indicating that larger degree is correlated with earlier age of acquisition.

Word class	Effect on AoA in ΔR^2				
	Log-AI	Log-WF	Log-AI after log-WF	Log-WF after log-AI	Log-WF + log-AI
All	0.11***	0.03***	0.09***	0.01*	0.12***
Nouns	0.08***	0.40***	0.01*	0.32***	0.41***
Verbs	0.14***	0.12***	0.06*	0.03	0.18***
Adjectives	0.13**	0.14**	0.04	0.04	0.18**
Function	0.20***	0.21***	0.18***	0.18***	0.39***

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 3
Results of linear regression of log-transformed CHILDES degree (i.e., log-CHd) on age of acquisition. Items in parenthesis are not log-transformed for CHd. In all cases the regression coefficients are negative, indicating that larger degree is correlated with earlier age of acquisition.

Word class	Effect on AoA in ΔR^2			Correlation Log-CHd and log-AI
	Log-CHd	Log-CHd after log-WF	Log-CHd + log-WF	
All	0.02*** (0.00)	0.07*** (0.11***)	0.10*** (0.14***)	0.24*** (−0.01)
Nouns	0.38*** (0.24***)	0.00 (0.02**)	0.40*** (0.42***)	0.61*** (0.512***)
Verbs	0.09** (0.04*)	0.12*** (0.05*)	0.24*** (0.18***)	0.54***(0.32**)
Adjectives	0.11* (0.09*)	0.12** (0.00)	0.26** (0.14*)	0.53***(0.37**)
Function	0.11** (0.19**)	0.14** (0.00)	0.35*** (0.21**)	−0.07 (0.09)

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 4
Results of the linear regression of log-transformed CHILDES degree (log-CHd) and log-transformed associative indegree on age of acquisition, after controlling for other factors. Items in parenthesis are not log-transformed for CHd. In all cases the regression coefficients are negative, indicating that larger degree is correlated with earlier age of acquisition.

Word class	Effect on AoA in ΔR^2	
	Log-AI after log-CHd	Log-CHd after log-AI
All	0.08*** (0.10***)	0.00 (0.00)
Nouns	0.01* (0.00)	0.31*** (0.16***)
Verbs	0.07** (0.11**)	0.01 (0.01)
Adjectives	0.05 (0.08*)	0.02 (0.03)
Function	0.23*** (0.17***)	0.14** (0.16**)

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

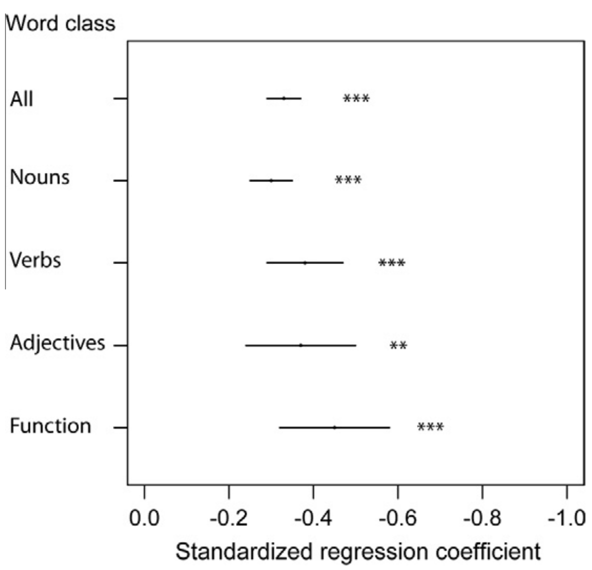

Fig. 2. Standardized regression coefficients for the different word classes, using the log of the associative indegree (log-AI) to predict age of acquisition. The negative regression coefficients indicate that as words have a higher log-AI, they are learned at an earlier age. Error bars are SEM. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Table 3 provides the results for a regression analysis using CHd with a window size of 5 to predict AoA. We report results for each word class, predicting AoA with CHd alone (column 1), CHd after controlling for word frequency (column 2), and both CHd and word frequency simultaneously (column 3). The results indicate that CHd is a predictor of AoA, and that this is true even after controlling for word frequency. Fig. 4 presents the standardized regression coefficients, showing that in all cases, higher log-CHd is associated with earlier AoA. The strongest effect is seen for nouns, and the weakest effect is found when all words are combined. The observation that the fit is higher within word classes than between them (All, column 1), suggests that, like frequency (see Goodman et al., 2008), contextual diversity may play different roles among different word classes. Further, the variance in AoA that is explained by contextual diversity combined with frequency (Table 3, third column) is greater overall than associative indegree (cf. Table 2, column 5). This is true of all word classes except function words, though not true over all words when combined. Log-transformed nouns are possibly another case where this may not be true. The sign of the coefficients are always negative, with more contextually diverse nouns being learned at younger ages. Finally, we also provide the correlation between a word's diversity in CHILDES and its associative indegree (Table 3, column 4). These are highly correlated for all word classes except function words. Adult free associates thus reflect (in part) the contextual diversity of the learning environment and their ability to predict age of acquisition of early-learned words may be due to that relationship.

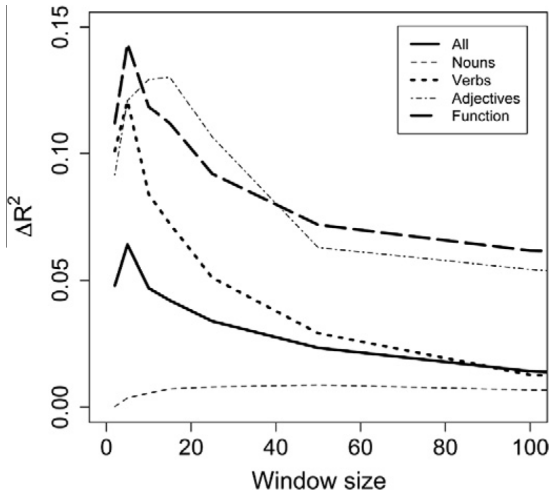

Fig. 3. Change in R^2 resulting from a linear regression of CHd on AoA. Lines represent all word types combined and the four different word types separately, after controlling for the log of the word frequency.

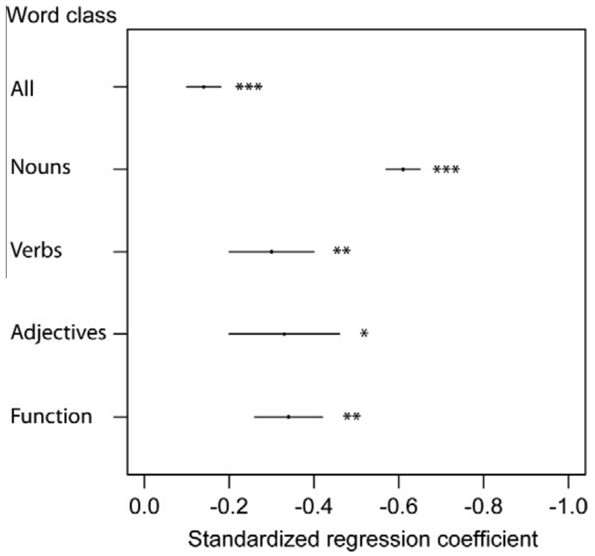

Fig. 4. Standardized regression coefficients for different word classes, using the log of CHILDES degree (log-CHd) to predict age of acquisition (column 1 of Table 3). The negative regression coefficients indicate that as words have a higher log-CHd, they are learned at an earlier age. Error bars are SEM. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

We also examined how much of the variance in AoA is explained by CHd after controlling for associative indegree. If contextual diversity is the more relevant factor for AoA, then most of the variance explained by associative indegree should disappear if we first take out the contribution made by CHd. Column 1 of Table 4 shows that this is only dominantly true for nouns, and possibly for adjectives—both of which show a large reduction in the variance explained by associative indegree after controlling for CHd. This means that, for these other word classes, associative indegree is indexing some other factors relevant to word learning that are distinct from contextual diversity as defined here. Also, as shown in column 2 of Table 2, after removing the effect of associative indegree, CHd is only a significant predictor of AoA for nouns and function words. Again, the statistical properties relevant for learning different kinds of words are shown to be different.

The above analyses suggest that contextual diversity, as measured by the count of unique other words a word co-occurs with in child-directed speech, is a significant predictor of age of acquisition for nouns, verbs, adjectives, and function words. The results also indicate a significant relationship with associative indegree—with strong positive correlations between associative indegree and CHd for all word classes but function words. This is consistent with our claim for an associative structure in caregiver speech, i.e., the idea that contextual diversity mediates the effect of associative indegree (especially for nouns). However, the results also support an *independent* effect of associates—one that is not reducible to contextual diversity, except in the case of nouns.

In summary, the predictive relation between adult free associates, nouns, and age of acquisition found in our previous work (Hills et al., 2009a) generalizes to other word classes, but differentially so depending on the word class. Adult associates are also correlated with contextual diversity in the language-learning environment. In the case of nouns, contextual diversity and adult associates are not independent in their prediction of age of acquisition. However, for other word classes, contextual diversity and adult associates do provide independent information relevant to age of acquisition.

With respect to acquisition mechanisms, these results are consistent with the hypothesis that some word classes are learned based on the contextual nature of the learning environment. If this is true, then it supports the theoretical and empirical positions that find evidence for a role of contextual diversity (as outlined in the introduction, e.g., Gillette et al., 1999; Hayes & Clark, 1970; Hurtado et al., 2008), but not those that support greater consistency. Moreover, it should also support a learning process via preferential acquisition, in which the contextual diversity of a word in the learning environment enhances that word's learning potential. If such learning is further enhanced when new words appear in contextually diverse backgrounds with *known* words, then the lure of the associates should also perform well. These enhancements should be most effective for nouns, but to a lesser degree for other word classes (see Fig. 4 and Table 3 column 1). We test these hypotheses in the following section.

Study 2: analyses of models of longitudinal network growth using contextual diversity

The three principle network growth models—preferential attachment, preferential acquisition, and the lure of the associates—offer different hypotheses about how children use statistical information both in terms of the structure of the learning environment and the statistical structure of the words they already know. As described in Fig. 1, preferential attachment is based on the connectivity of known words; preferential acquisition is based on the connectivity of new words to all words in the learning environment; and the lure of associates is based on the connectivity of new words to known words. In the preferential attachment model, a word is more likely to be learned if it attaches to an existing already *known* word in the network that is itself well attached. In this way, richly connected words become more richly connected. This pattern of growth has been proposed for a number of real-world systems, including the Internet, highways, and for the development of protein-interaction networks (e.g., Albert & Barabasi, 2002; Pastor-Satorras, Smith, & Sole, 2003). Steyvers and Tenenbaum (2005) also hypothesized that this pattern of growth – one in which a word is more likely to be learned if it attaches to a well connected word already in the growing network– characterizes lexical development. With preferential acquisition, on the other hand, a word is more likely to be learned if it is attached to many other words *in the learning environment*. The lure of the associates lies between the preferential attachment and preferential acquisition models: at the time of acquisition, the child learns the word that attaches to the most already known words.

Our previous work indicates that when adult associates are used to connect nouns in a developing network, preferential acquisition is the better fitting growth principle (Hills et al., 2009a). Here we extend these analyses by building networks of all word classes, not just nouns, and building them based on the co-occurrence statistics in the learning environment.

Methods

Networks

As in our previous work, developing networks were built by selecting subsets of words from the CDI that are normatively known at different ages. We then built co-occurrence networks for each subset, with each word pair being connected by a directed link if one word co-occurred before another word in the language-learning environment (cell $ij > 0$), as indicated by the CHILDES corpus (represented by arrows in Fig. 1). This allowed us to create 15 networks based on contextual diversity, corresponding to one-month increments in the CDI data. Therefore, each month's network represented the words that were normatively acquired by that month, connected by their co-occurrence relationships in the learning environment. For example, the 16-month co-occurrence network contained 17 words – since 17 words were normatively acquired by 50% of children by 16 months – and links representing whether or not these words co-occurred in the corpus of caregiver speech. This set of 15 networks was then used to test the three growth models, as described below, by determining the structural properties of the new words added at each month. The growth model of preferential acquisition requires that we describe the structure of the learning environment, which includes words not yet known by the language learners. Accordingly, we chose two representative learning environments to test preferential acquisition. One, the 30-month network, consisted of only words specific to a given word class in the CDI, known by more than 50% of children at 30 months of age. The other, the "adult" network, used the combined 532 words from all word classes plus the 4468 most frequent words in the CHILDES corpus.

Lexical growth is likely to depend on factors other than semantic relatedness or contextual diversity. Accordingly, the analyses also considered properties such as phonotactic probability and the number of phonological neighbors since these have also been shown to influence the potential acquisition of early-acquired words (Charles-Luce & Luce, 1995; Landauer & Streeter, 1973; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985; Storkel, 2001). Indeed, in previous work, we found an effect of the number of phonological neighbors on the learning of nouns as well as associative connections to other nouns, with nouns having more phonological neighbors being learned earlier (Hills et al., 2009a).

Statistical analysis of growth models

This analysis considered words connected by co-occurrence; that is CHd. For each month, the probability that a word w_i is added to the network is based on its learning value, d_i:

$$P(w_i) = \frac{e^{\beta d_i}}{\sum_j e^{\beta d_j}} \quad (1)$$

The weight for new words was determined by a parameter β, which represents the sensitivity of the acquisition process to d_i. In particular, positive values of β mean that words with higher values of d_i were more likely to be acquired early (having a higher probability of being learned), whereas negative values of β mean that words with low values of d_i were more likely to be acquired early. We let d_i represent the learning values ("Value" in Fig. 1) for each word calculated with respect to each model. For example, with the lure of associates model and at a specific month, d_i is equivalent to the degree of the new word i if it were added to the known network at that month. The denominator was calculated for all words that were not yet learned at the start of the month for which the word in the numerator is acquired. The log of the $P(w_i)$ values, for each acquired word, was then added to produce the log likelihood, as follows:

$$L(\beta) = \sum_i \log(P(w_i)) \quad (2)$$

We then found the β^* that maximized the above log likelihood function using a standard optimization procedure. To compare models, we then computed the G^2 for each model, M, as follows:

$$G^2 = 2(L_M - L_{\text{random}}) \quad (3)$$

which is a measure of the models improvement in log likelihood value when compared with a random learning model (Burnham & Anderson, 1998; Busemeyer & Diederich, 2010). Our random model assumed that all words learned up to 30 months are equal in their learning value. To avoid overfitting to words in the later months, the analyses only used words learned in the first ten months (up to month 26). We then compared models using the Bayesian Information Criterion (BIC; Schwarz, 1978). BIC penalizes models according to their complexity, with k additional parameters and N observations, according to the following equation:

$$\text{BIC} = G^2 - k \cdot \ln(N). \quad (4)$$

Larger values of BIC indicate models that are a significant improvement over lower valued BIC models.

The analyses also compared a number of additional growth factors that may influence order of acquisition, including the number of phonemes in a word, the number of phonological neighbors, average phonotactic probability, and frequency in the CHILDES corpus (values were taken from Balota et al., 2007; Vitevitch & Luce, 2004). For example, to evaluate the role of phonological neighbors, we assigned the number of phonological neighbors for a word to its learning value in Eq. (1). When combining possible growth factors (e.g., preferential acquisition using CHd and phonological neighbors), we additively combined the product of individual β for each growth factor, v, and its associated learning value in the exponent of Eq. (1), as follows:

$$P(w_i) = \frac{e^{\sum_v \beta_v d_{i,v}}}{\sum_j e^{\sum_v \beta_v d_{j,v}}} \quad (5)$$

Results and discussion

Table 5 presents the results for the model comparison using contextual diversity in the CHILDES corpus. The

Table 5
Bayesian Information Criterion for the three growth models. The upper portion of the table shows the models, when tested individually. The middle upper portion shows the results for frequency and number of phonological neighbors, when tested individually. The middle lower portion of the table presents the best individual model from above (shown in bold) with additional variables representing the frequency of the word in the CHILDES corpus (Freq. CHILDES) both alone and with an additional variable representing the number of phonological neighbors (Phono. Neigh.). The lower portion of the table provides the number of possible words that could be learned (the total minus those already known in the first month) and the number of words that were actually learned over the period of analyses up to month 26. Bold numbers indicate the best growth model of those presented in Fig. 1.

Model	Word class				
	All	Noun	Verb	Adjective	Function
Preferential attachment	−5.61	0.35	−2.74	−2.64	−0.05
Lure of the associates	**3.42**	**76.87**	0.10	−3.24	1.13
Preferential acquisition – 30 months	0.31	76.09	**0.90**	−2.01	10.19
Preferential acquisition – adult	−2.18	67.40	−0.07	−**1.16**	**19.64**
Frequency in CHILDES	−0.37	68.92	0.97	0.06	20.16
Phonological neighbors	−3.83	8.49	−3.95	−1.94	−1.25
Best growth model+					
Freq. CHILDES	−1.00	71.82	−3.18	4.39	17.45
Freq. CHILDES and Phono. Neigh.	−0.73	66.30	−6.43	5.05	14.77
Number of words learned	395	250	72	37	15
Number of possible words	516	299	91	57	50

upper portion presents the model BIC values individually for each model and for each word class. The results indicate that for all word classes combined and nouns, the lure of the associates is the best fitting model. For function words and verbs, the best fitting model is preferential acquisition. And for adjectives, none of the models have a positive BIC. In all cases, the model sensitivity parameters, the βs in Eq. (1), are positive and nonzero, indicating that words with larger learning values with respect to each model are more likely to be added to the semantic network earlier during development. This is consistent with our previous results (Hills et al., 2009a), showing that preferential attachment performs worse than the other growth models for nouns. Here we find that it also performs poorly for other word classes. Across word classes, lure of the associates and preferential acquisition perform best, with lure of the associates outperforming preferential acquisition when the analysis includes all words or when it includes only nouns.

None of the growth models individually show a significant contribution to the learning of adjectives. Recall that the best window size for adjectives was 15. When using this window size, preferential attachment significantly outperforms the other models (the rest of which are no different from the random model).

In the middle portion of Table 5, we also see that frequency in CHILDES is a significant predictor of order of acquisition for all word classes, but performs best for nouns. Number of phonological neighbors only predicts order of acquisition for nouns. In the middle lower portion of the table we combine frequency in CHILDES and the number of phonological neighbors with the best model from the upper portion of the table (shown in bold). Only in the case of adjectives does this significantly improve the prediction of order of acquisition over the individual models when tested alone.

In summary, the model analysis supports the notion that different word classes are learned differently. While none of the word classes support preferential attachment, all the word classes (except adjectives) support a learning process that is sensitive to the structure of words in the learning environment (preferential acquisition), and also how these words co-occur with words that are already known (the lure of the associates).

General discussion

The analyses reported in this paper make five contributions. First, they provide insight into the observed relation between adult associations and the age of acquisition of early-learned words, with some indication that this effect is partially mediated by contextual diversity, especially for nouns. Second, the analyses provide evidence for the processes that underlie the early growth of semantic networks, supporting the growth patterns of preferential acquisition and the lure of the associates. Third, they show that contextual diversity is positively related to lexical development, a result relevant to understanding why individual words are acquired earlier than other words. Fourth, the results show that at least in aggregate the positive effect of contextual diversity on learning holds across different word classes, including nouns, verbs, adjectives, and function words. Fifth, the results indicate the statistical properties that matter for lexical learning may be different for different classes of words. We consider these contributions in turn and then their limitations and next steps.

Adult associations and age of acquisition

Adult free associations are a strong index of semantic relatedness and have been shown to be robust in predicting many forms of adult semantic judgments (e.g., Nelson, Zhang, & McKinney, 2001). In summarizing a large body of work, Deese (1965) concluded that such word associations reflect the contiguity, semantic, and frequency properties of words in the language. Statistical properties related to co-occurrence appear to be particularly critical to these associations; words that appear together in language more frequently also have a higher likelihood of appearing in associative pairs (Lund, Burgess, & Audet, 1996; Spence & Owens, 1990). More recently, Steyvers and Tenenbaum (2005) used word associations to construct networks of semantic relatedness of large numbers of words such that words with common associates were linked in the network. These semantic networks displayed a power-law distribution of links over nodes; the network contained a few hubs (i.e., a few high-degree nodes with many links to other nodes) and many nodes with only a few connections. They also showed that connectivity in these networks was correlated with age of acquisition, a result also found by Hills et al. (2009a).

The present results show that adult associations are also related to contextual diversity in the learning environment, which has important consequences for interpreting prior research. It raises the possibility that contextual diversity is at least *part* of the reason associates are correlated with age of acquisition. It further suggests that the age of acquisition effect—related to adult word recognition, lexical decision and word retrieval times (e.g., McEvoy, Nelson, & Komatsu, 1999; Steyvers et al., 2005)—may be in part driven by contextual diversity.

Adult associations, contextual diversity, and the learning environment all reflect the structure of language itself as well as language use (Lund et al., 1996; Spence & Owens, 1990). Consistent with this connection, contextual diversity is related to lexical decision tasks (Adelman et al., 2006; Hicks, Marsh, & Cook, 2005; Recchia et al., 2008; Steyvers & Malmberg, 2003), and age of acquisition is related to word recognition and production in adults (Ellis & Morrison, 1998; Stewart & Ellis, 2008). Because contextual diversity is itself driven by the nature of language and the world, we cannot know for certain that contextual diversity (and not some other structural property to which it is related) is the key factor in language acquisition or lexical decision tasks. Still, the entire pattern suggests a reasonable hypothesis: Early-learned words that co-occur with many other words in the language-learning environment will as a consequence be densely associated to many other words, forming connected hubs in the semantic network. As a result, these early-learned words are the words most likely to be activated in any given context. This may benefit learning directly. It may also benefit lexical access and retrieval and so lead to the structure of adult associations. Although presently underdetermined, this unifying hypothesis is empirically testable both by measuring word lexical access in young children and through artificial-language learning studies.

Processes of network growth

The power-law distribution of links over nodes, found previously for adult networks based on free associates, can be shown to emerge by a process called preferential attachment (Barabasi & Albert, 1999; Pastor-Satorras et al., 2003; Steyvers & Tenenbaum, 2005). While Steyvers and Tenenbaum (2005) hypothesized this pattern of growth for lexical development, Hills et al. (2009a) showed that two alternative patterns of growth – what they called preferential acquisition and lure of associates – could better account for the emergence of a power-law distribution in developing semantic networks. The key difference between preferential attachment and the other two models of network growth is that growth by preferential attachment depends only the structure of the growing network whereas both preferential acquisition and lure of associates depend on the structure of the learning environment and its relationship to what is known, respectively. More specifically, preferential acquisition depends only on the structure of the learning environment, while the lure of the associates focuses on the interface between what is known and what can be known. These would seem to be potentially relevant differences between lexical development and the growth of the Internet. Though, of course, the Internet too may grow by a process of preferential acquisition or the lure of the associates—if we allow that there is an information structure in the world prior to its appearance on the Internet.

The present results provide further support that it is the statistical structure of the learning environment and not just (or even) the emerging structure of the semantic network that creates the relation between patterns of connectivity and age of acquisition for individual words. The present results extend the results of work like Hayes and Clark (1970) by showing that early word learning is associated with contextual diversity using naturally collected child-directed speech over a period of months. In so doing, we provide further evidence for a proximal mechanism in the early acquisition of words that *come to be* hubs in adult semantic networks. The proximal cause appears to be contextual diversity; highly diverse words are more easily disambiguated—i.e., more readily isolated from the background—and perhaps also more easily mapped to their corresponding referents. As Yu and Smith (2007) demonstrated in cross-situational word learning with adults, greater contextual diversity – given cluttered and ambiguous learning events – leads across those events to fewer spurious correlations and thus better learning.

Contextual diversity and early vocabulary growth

Parents' conversations with their children provide the data on which lexical development depends. Accordingly, many researchers in child language have sought to understand the relationship between parental speech and vocabulary growth. Many of these studies have examined correlations between the structure of individual parents' speech and the structure of vocabulary growth in those parents' children (see Hart, 2004; Hoff & Naigles, 2002; Huttenlocher et al., 1991). One consistent finding of these studies is a strong relationship between lexical richness in the input, that is the number of different words, and the lexical richness and rate of the vocabulary growth in the child (see especially, Hoff & Naigles, 2002). On the one hand, this is not at all surprising: after all, a child can only learn the words they hear. A parent who says many different kinds of words is likely to have a child who learns many different words, whereas a parent who repeats the same words – with little diversity – is likely to have a child who knows and uses few words. On the other hand, this correlation in parent-child lexical diversity may be fundamentally important for understanding lexical acquisition: as many researchers of child language have noted, linguistic context matters and hearing individual words in a variety of linguistic contexts should help the learning of those words (Gleitman, 1990; Hart, 2004; Naigles, 1996; Naigles & Hoff-Ginsberg, 1998; Waxman & Markow, 1998). Lexical diversity in the input should not be relevant just to explaining individual differences in the range of vocabulary growth between children, but should also be relevant to explaining why some words are learned earlier than other words. The present analyses show that contextual diversity – which is necessarily related to the diversity of words in the learning environment as a whole – explains at least some of the variance in age of acquisition and does so above and beyond word frequency and phonotactic properties. A key next step for a more detailed understanding of the relationship among (1) lexical and contextual diversity in the input, (2) individual differences and early vocabulary growth, and (3) age of acquisition of individual words will be mapping the statistical structure of individual parents' speech to the lexical development of their children.

The results presented here may also have implications for learning beyond early development. Given the logical problem of indeterminancy set down by Quine (1960), any contextual cues that provide conceptual boundaries for words also provide a means for better understanding those words and parsing the words from the speech stream—regardless of the learner's age. Medin and Ross (1989) point out that when examples of a phenomenon are too similar, individuals are likely to make 'conservative generalizations' that include many of the irrelevant background features; diverse examples better resolve these conceptual boundaries. Thus the advantage of contextual diversity may help resolve both phonetic word boundaries as well as conceptual word boundaries, and may do so for both adults and children.

Nouns, verbs, and adjectives

The results show that the predictive power of contextual diversity is effective for all word classes, but differentially so. Nouns show the strongest effect of contextual diversity in child-directed speech, but much of this effect is removed after controlling for frequency (Table 3, column 2). However, for all other word classes, the effect of contextual diversity improves after controlling for frequency (Table 3, column 2). The pattern observed here is consistent with growing evidence that the frequency of words in child-directed speech is related in complicated ways to acquisition and differs for different word classes (Goodman et al., 2008; Sandhofer et al., 2000). In brief, there is no *overall* positive correlation with the frequency of words and age of acquisition. A recent analysis by Goodman et al. (2008) showed that there are, nonetheless, strong frequency correlations *within* a word class (e.g., more frequent nouns are learned earlier than less frequent nouns) but the relation between frequency and class of words is just the opposite. Function words such as articles and prepositions are very frequent and learned relatively late; individual basic-level nouns—given the sheer size of this open class set of words—are individually relatively infrequent but are early acquisitions. Goodman et al. (2008) concluded that frequency effects on age of acquisition of words are nonlinear and complicated. It seems likely that no single statistical property—e.g., frequency or contextual diversity—will simply or straightforwardly predict age of acquisition for all word classes simultaneously. Our findings (consistent with others, e.g., Recchia et al., 2008) make a clear prediction that contextual diversity is at least as important as frequency, especially for nouns—where we saw the largest effect of their combined predictive power (see Table 3). This noun difference may be due to the fact that learning nouns is primarily about matching words with objects in the environment, which is facilitated by the interaction of frequency and diversity (e.g., Yu & Smith, 2007). Other word classes may have still more complicated pathways for learning, and more research will be needed to resolve what those are.

It is interesting to note that preferential acquisition and its close neighbor, lure of associates, were the best fitting growth models for all word classes *but* adjectives. For adjectives (descriptive words on the CDI) no model was a clear winner using a window size of 5. However, with a window of size 15 (where adjectives peaked in Fig. 2), preferential attachment was the best model for adjectives. If this better fit by preferential attachment holds up under further investigation, this suggests that the structure of the adjectives already known is a better predictor of what new adjectives will be learned than the structure of the learning environment. Backscheider & Shatz (1993; see also, Tare, Shatz, & Gilbertson, 2008) and Sandhofer and Smith (1999) came to the same conclusion in their analyses of children's learning of color terms.

The finding that contextual diversity predicts age of acquisition for early-learned verbs fits Hoff and Naigles (2002) findings that the diversity in the syntactic context in which individual verbs are heard predicts the diversity in children's use. Our findings lend support to the general conclusion that for verbs, like nouns, hearing to-be-learned words in a variety of contexts helps learning. This resolves the question of the benefit of early consistency versus diversity for verb learning in the favor of diversity. However, the arguments and evidence for the benefit of hearing and use of verbs in narrow syntactic and semantic contexts primarily concern the earliest stages up for learning (see Akhtar & Tomasello, 1997; Goldberg, 2006; Goldberg et al., 2004; Sethuraman, 2004; Tomasello, 2000). The present analyses did not include an examination of possible changes in the statistical structure of child-directed speech as a function of the developmental level of the child. Sethuraman and Goodman (2004; see also Goldberg, 2006) reported evidence suggesting that consistency – i.e., a narrow range of contexts – for verbs in the input at 18 months predicts better acquisition of those verbs at 24 months. Therefore, it will be important (though difficult given the reduction in size of the available corpora to be analyzed) to investigate the statistical structure of child-directed speech at different points in lexical development.

Limitations and open questions

Contextual diversity is both a fact about the learning environment and a product of whatever it is that makes people talk the way do. Thus, it is possible that at the level of word learning mechanisms, it is not contextual diversity that matters per se, but other processes that cause people to talk about what they do. For example, things young children care about may be talked about in more contexts (because children care about them) and learned, not because they are talked about in more contexts, but because children care about them. Analyses of the present sort – descriptions of the learning environment and structural descriptions of the order in which words are learned—do not address the question of learning mechanisms and so in and of themselves cannot address this issue. However, the results do point to one aspect of the learning environment (contextual diversity) that matters to the structure of learning (the order in which words are learned) and in so doing provides a link between learning history (the order in which words are learned) and adult's performance (free associations). This seems a promising step toward defining the data that need to be explained. Further, although the correlational patterns reported here cannot show that contextual diversity itself matters to the word learning process, this statistical property has been shown to be relevant to many candidate learning mechanisms (e.g., Glenberg, 1979; Hayes & Clark, 1970; Rogers & McClelland, 2003; Verkoeijen et al., 2004).

In sum, all the evidence reported in this paper is correlational and thus provides circumstantial but not causal evidence for a role of contextual diversity in acquisition or for the relation between age of acquisition and adult lexical processing. For example, the age of acquisition effect—in which lexical decision times are related to the age of a word's acquisition—might be interpreted as a situation where early acquisition in some way sets up processes that yield more rapid lexical retrieval in adults (Ellis & Morrison, 1998). Alternatively, contextual diversity –and the network structure it engenders – could be the root cause of both age of acquisition and adult retrieval times. In the final analysis, it seems likely the direction of causation goes in more than one direction and through more than one mediating process (e.g., Recchia et al., 2008; Stewart & Ellis, 2008). Word learning (and word retrieval) are most likely a self-reinforcing dynamical system, in which the earliest learned words become more easily retrieved during speech, and thus reinforce the learning of these words earliest in future generations through a process involving contextual diversity (see Gershkoff-Stowe & Smith, 1997; Gershkoff-Stowe & Hahn, 2007, Gershkoff-Stowe, 2002, for relevant empirical evidence). The structure of language influences both real time processes of language learning and language use, and over time these create not only the long-term stable knowledge structures that support language learning and language use, but also the language-learning environment for the next generation of learners.

Acknowledgments

This work was supported by the National Institute of Health, T32 Grant HD 07475 and by NIH/NICHHD R01HD28675 to Linda Smith.

References

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review, 115*, 214–229.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.

Albert, R., & Barabasi, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*, 47–97.

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952–965.

Backscheider, A. G., & Shatz, M. (1993). Children's acquisition of the lexical domain of color. In K. Beals et al. (Eds.). *What we think, what we mean, and how we say it* (Vol. 2, pp. 11–21). Chicago: The Chicago Linguistic Society.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509–512.

Brown, P. (2007). 'She had just cut/broken off her head': Cutting and breaking verbs in Tzeltal. *Cognitive Linguistics, 18*, 319–330.

Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference. A practical information-theoretical approach*. New York: Springer-Verlag.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Los Angeles: Sage Publications, Inc.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighborhoods in young children's lexicons. *Journal of Child Language, 17*, 205–215.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers, 28*, 125–127.

Deese, J. (1965). *The structure of associations in language and thought*. Baltimore: Johns Hopkins University Press.

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 515–523.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49*, 988–998.

Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language, 46*, 665–687.

Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research, 50*, 682–697.

Gershkoff-Stowe, L., & Smith, L. B. (1997). A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology, 34*, 37–71.

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition, 7*, 214–223.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition: A Journal of Developmental Linguistics, 1*, 3–55.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*, 95–112.

Goldberg, A., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*, 289–316.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldin-Meadow, S., Goodrich, W., Sauer, E., & Iverson, J. (2007). Young children use their hands to tell their mothers what to say. *Developmental Science, 10*, 778–785.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*, 515–531.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science, 18*, 1069–1076.

Hart, B. (2004). What toddlers talk about. *First Language, 24*, 91–106.

Hayes, J. R., & Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An observation on the role of contextual variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1160–1164.

Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20*, 729–739.

Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Categorical structure among shared features in networks of early-learned nouns. *Cognition, 112*, 381–396.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*, 418–433.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*, F31–F39.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236–248.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Language and Cognitive Processes, 16*, 565–581.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks, 17*, 1345–1362.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavioral Research Methods, Instruments, and Computers, 28*, 203–208.

Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. In *Proceedings of the Cognitive Science Society* (pp. 603–608). Hillsdale, NJ: Erlbaum.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (Third Edition). Mahwah, NJ: Lawrence Erlbaum Associates.

Maouene, J., Laakso, A., & Smith, L. B. (submitted for publication). Object associations of early-learned "light" and "heavy" English verbs.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: The MIT Press.

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What's the connection between true and false memories: The different roles of inter-item associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177–1194.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, Instruments, and Computers, 37*, 547–559.

Medin, D. L., & Ross, B. H. (1989). The specific character of abstract thought: Categorization, problem-solving, and induction. In R. J. Sternberg (Ed.). *Advances in the psychology of human intelligence* (Vol. 5, pp. 189–223). Hillsdale, NJ: Erlbaum.

Meints, K., Plunkett, K., & Harris, P. (2008). Eating apples and house plants: Typicality constraints on thematic roles in early verb learning. *Language and Cognitive Processes, 23*, 434–463.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Emory cognition project conference. The ecological and intellectual bases of categorization, Oct 1984, Atlanta, GA, US* (pp. 201–233). New York, NY, US: Cambridge University Press.

Mervis, C. B., Mervis, C. A., Johnson, K. E., & Bertrand, J. (1992). Studying early lexical development: The value of the systematic diary method. *Advances in Infancy Research, 7*, 291–378.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review, 111*, 721–756.

Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition, 58*, 221–251.

Naigles, L. R., & 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, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://w3.usf.edu/FreeAssociation>.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review, 99*, 322–348.

Nelson, D. L., Zhang, N., & McKinney, V. M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1147–1159.

Newman, R. (2008). The level of detail in infants' word learning. *Current Directions in Psychological Science, 17*, 229–232.

Pastor-Satorras, R., Smith, E., & Sole, R. (2003). Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology, 222*, 199–210.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*(1), 161–167.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75–95.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *Carnegie Mellon symposium on cognition, May 1997, Pittsburgh, PA, US* (pp. 381–415). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

R Development Core Team. (2009). R: A language and evnironment for statistical computing. R Foudnation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.

Recchia, G., Johns, B. T., & Jones, M. N. (2008). Context repetition benefits are dependent on context redundancy. In *Proceedings of the 30th Annual Cognitive Science Society*, Washington, DC, USA, July 23–26.

Riordan, B., & Jones, M. (2007). Comparing semantic space models using child-directed speech. In D. S. MacNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society*.

Rogers, T., & McClelland, J. (2003). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35*, 185–205.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Sandhofer, C. M., & Doumas, L. A. A. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development, 9*, 194–221.

Sandhofer, C. M., & Smith, L. B. (1999). Learning color words involves learning a system of mappings. *Developmental Psychology, 35*, 668–679.

Sandhofer, C., & Smith, L. B. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development, 3*, 233–267.

Sandhofer, C. M., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: Differential frequencies, different kinds of learning? *Journal of Child Language, 27*, 561–585.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 5*, 461–464.

Sethuraman, N. (2004). Influence of parental input on learning argument structure constructions. In *On-line proceedings supplement of Boston University child language development 28*. <http://www.bu.edu/linguistics/APPLIED/BUCLD/supp.html>.

Sethuraman, N. & Goodman, J. C. (2004). Children's mastery of the transitive construction. *Paper presented at the Stanford Child Language Research Forum*, April 2004.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research, 19*, 317–330.

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age of acquisition effect? *Psychonomic Bulletin and Review, 15*, 70–74.

Steyvers, M., & Malmberg, K. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*, 760–766.

Steyvers, M., Shiffirin, R. M., & Nelson, D. L. (2005). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 237–249). Washington, DC: American Psychological Association.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*, 41–78.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321–1337.

Tare, M., Shatz, M., & Gilbertson, L. (2008). Maternal uses of non-object terms in child-directed speech: Color, number and time. *First Language, 28*, 87–100.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156–163.

Verkoeijen, P. P., Rikers, R. M., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 796–800.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*, 481–487.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*, 571–581.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.

Waxman, S. R., & Markow, D. B. (1998). Object properties and object kind: Twenty-one-month-old infants' extension of novel adjectives. *Child Development, 69*, 1313–1329.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.