Title	Different Classes of Words are Learned in Different Ways
Author(s)	Hidaka, Shohei
Citation	Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society (CogSci 2009): 2316-2321
Issue Date	2009
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/9537
Rights	Copyright © 2009 Shohei Hidaka. Different Classes of Words are Learned in Different Ways, Shohei Hidaka, Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society (CogSci 2009), 2009, 2316-2321.
Description	

Different Classes of Words Are Learned in Different Ways

Shohei Hidaka (shhidaka@indiana.edu)
Department of Psychology and Brain Sciences
1101 E. Tenth Street, Bloomington, IN 47405-7007 USA

Abstract

What determines vocabulary growth patterns? The research presented here examines the growth pattern of words listed in the McArthur-Bates Communicative Development Inventory using a computational model. Our model characterizes vocabulary growth curves based on the sampling of learning relevant events and a threshold (the number of such events needed) for acquisition of the word. Using this general class of models, fits of vocabulary growth curves suggests a transition from one in which acquisition is primarily limited by the threshold for acquisition to one in which acquisition is primarily limited by sampling speed. Further analyses suggest that these parameters of the learning model link to meaningful psychological factors: specifically the acquisition of threshold limited (and earlier learned) words are correlated with frequency whereas sampling-speed-limited words are correlated with imageability of the word in the input.

Keywords: Vocabulary Growth; Age of Acquisition; Statistical Analysis.

What predicts vocabulary growth patterns?

In the first years of life, children begin to comprehend and produce words. Between 8 and 16 months of age, children's receptive vocabularies nearly double in size every two months (Dale & Fenson, 1996). From 12 to 24 months, their expressive vocabularies follow a similar path of productive growth. It has been estimated that between 18-months and 18-years of age children acquire approximately ten new words per day, or one new word every hour and a half the child is awake (Bloom, 2000). The words a child learns in this time period include nouns, verbs, determiners, preposition, however, nouns are acquired at a faster rate than other word classes.

What processes underlie this efficient learning pattern? Traditionally the growth pattern was described in terms of a sudden acceleration proposed to occur when children are 18-month-old or their number of acquired words reaches 50 words. This idea of a vocabulary spurt suggests a unitary change the sudden realization that things have names (Reznick & Goldfield, 1992), the onset of categorization abilities (Gopnik &Meltzoff, 1987), or the acquisition of word learning constraints (Mervis & Bertrand, 1994). More recent accounts conceptualize the process in terms of a single or set of self-accelerating processes (van Geert, 1998). This view recognizes the fact that the age of acquisition (AoA) of any word will depend on a variety of factors: frequency, word length, phonological similarity, semantic similarity, lexical density, familiarity, imageability, and etc. Moreover, these variables tend to be correlated to each other. For example, more frequent words tend to be more familiar in general and to appear in more diverse contexts. All this makes the prediction of the age of acquisition of any single word complicated indeed. Also patterns of acquisition are subject to show considerable individual differences, which makes sense in many factors that are characteristic of individual experiences matter (see Bates, Dale, & Thal, 1995). One way to get a handle on all this is to consider vocabulary growth from a population perspective – populations of words with various properties in the learning environment and populations of children learning those words. This is the approach taken here. We examine the *growth patterns of the proportion of children* in large normative studies reported to "know" a word as an index of the properties of words and learners that make for earlier and later acquisition.

Our approach is based on the following ideas: (1) many complexly related psychological variables relevant to age of acquisition (2) generalization of growth patterns from individual children to population level patterns *and vice versa* requires an integration of these relevant factors *via a learning process* to outcome (AoA). Considering only the correlation between relevant factors in the input (e.g., frequency, diversity, concreteness) and output (AoA) is not enough to describe a coherent picture of word learning. Thus, the larger goal of this work is to specify the triplet, not a dyad, the relation among the properties of words, the learning process, and AoA. Here we present initial results that build on recent findings by Goodman et al. (2008) of nonlinear effects of frequency on AoA, a result which clearly suggests the complexity of the processes that determine the AoA of any word.

It's Complicated

There is general agreement that no simple psychological variable can account for the entire range of vocabulary growth. Although several studies show the some part of vocabulary growth seems strongly related to the input frequency of the words, the effect of frequency on vocabulary growth itself is also not simple. Goodman et al. (2008) analyzed the frequency effect on vocabulary growth in different classes of words and found that the AoA of words in the MCDI have a complicated correlational structure to frequency of words. Across the entire corpus of early learned words, there is a very small correlation between AoA and the frequency of the word in child directed speech. However, their analyses also suggest that this is a surface outcome of two different frequency effects. That is, they found that AoA and frequency of words *within classes* (e.g., with the noun class or within the verb class) are positively correlated (the more frequent, the earlier). However, the average AoA for a class and average frequency of words in a class (i.e., the *between class correlations*) were negatively correlated. For example, function words are highly frequent but learned late. Although frequency matters (the within-class effects), its relevance is complicated and clearly modulated by other factors.

Learning Process

How and why properties such as the frequency of words matter depend on the learning process, which may itself depend on the kind of word to be learned (see Sandhofer, Smith & Luo, 2000). There are various kinds of learning mechanisms to be considered that relate the input to the output, including connectionist and associative learning (Plunkett, 1993), Bayesian inferences (Xu & Tenenbaum, 2007), and so on. Within these models, it is possible to conceptualize a learning mechanism in which learning gradually accelerates; this could be due to generalization of statistical regularity in connectionism models, learning of higher-order level representation in Bayesian inference. And, indeed, there is empirical evidence from children that teaching them words does actually speed their acquisition of subsequent words (Gershkoff-Stowe & Smith, 2004).

Alternatively, the shape of the growth curve may reflect properties of a population of learners and a population of words; not the learning mechanism itself. For example, vocabulary growth curves (in the form of the number of words known by an individual child or the numbers of words known by a population of children as a function of age) are often well fit by logistic models (van Geert, 1998; Fenson et al., 2000). According to these logistic models, both word acquisition and learning rate depend on the threshold of a given variable (e.g., age, frequency or etc), but the underlying process (variable) goes from through sub- to super-threshold at a constant rate. A similar idea using a threshold but formulated in a slightly different way is proposed recently (McMurray, 2007).

Summing up this proposal, some approaches have focused on the learning processes and suggest there is an actual acceleration in the rate at which new words are acquired as a function of age and/or vocabulary size; other approaches have focused on the shape of the learning curve itself and suggest that there may be no underlying change in the rate of learning new words as vocabulary itself grows. Between these two dichotomous options (i.e., acceleration or constant speed) there are additional possibilities such as the shape of the learning curve and the nature of changes in that shape as function of vocabulary growth depending on the properties of the to-be-learned words.

The Approach in the Present Study

Our approach has three parts. First, we examine the growth of individual words (and classes of words) in terms of the proportion of children (in a large normative study) who are said to know that word at mo   nthly age intervals. Thus, we examine growth not in terms of the number of words and an individual child knows as a function of age (the usual sense of a growth curve) but in terms of a population of children. This is a potentially useful approach for thinking about the relation between the properties of words in the learning environment (which are, after all, population statistics rather than the statistics for individual children) and their likelihood of being learned early or late. More specifically, we use the month by month data from the MCDI data on the proportion of children from 16 to 30 months who are reported by parents to produce each of 654 early learned words (Fenson et al., 1993).

Second, we consider a general model of growth –one that includes the possibility of both an accelerating rate of acquisition and a constant learning rate. We ask how well specific cases of this model fit the data.

Finally, we consider how properties of different kinds of words relate to the observed results in terms of different growth patterns for different kinds of words. We specifically examine word frequency in adult use and in child-parent conversation, measure of familiarity, of imageability, of number of associations, etc.

The Model We propose a simple computational model which conceptualizes word learning in terms of a sampling of relevant events with the word being acquired once its passes a given threshold. The model assumes that children acquire a word if they are exposed to a given number of events relevant to acquisition of the word (McMurray, 2007). There are two theoretically important parameters – the threshold for acquisition and the learning rate. The model assumes the number of the sampled events is linearly or polynomially correlated to physical learning time (i.e., age, square of age, square root of age etc.). Thus, the model considers word learning as sampling until a given threshold is met and that learning may be constant, accelerated or decelerated in terms of the sampled events. More specifically, this is an extension of two statistical distributions such as the gamma and Weibull distribution. Since it is natural extension that includes these two models as special cases, we call it the hyper gamma-Weibull model. Note that this model provides a description of the shape of the learning curve and its key parameters do not specify in any straightforward way specific psychological mechanisms. Hypotheses about those mappings –from parameters of the growth curve to learning processes – are developed from the properties of the words that are best fit by different versions of this general model.

Any vocabulary growth curve is roughly S-shape, and models with a constant learning rate or an accelerated learning rate can generate such curves. However, under close examination different classes of curves can be distinguished. To help readers grasp the big idea, we show how different models yield somewhat different growth curves in Figure 1. The top row shows the cumulative density function the gamma, Weibull, hyper gamma-Weibull, and the logistic model. The bottom row shows their hazard functions which are conditional probabilistic densities. In other words, the hazard function may be interpreted as the probabilistic density that children who have not acquired a word at one moment will acquire the word at another given moment. All four distributions have the same mean value, and their cumulative distributions all look similar (i.e., S-shape). However, the hazard plots reflect the underlying different parameters in these models. The hazard function of gamma model is convex up, but that of Weibull model is convex down. The hyper model (given particular parameter settings) has a peak, and that of logistic model is identical to its cumulative density function except for the scale. Next we specifically describe how the computational process in the gamma-Weibull model is related to the vocabulary growth curve.

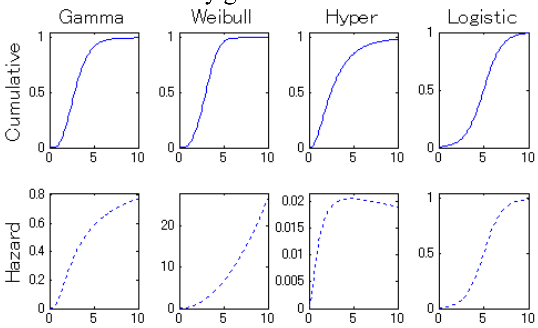

Figure 1: The cumulative density and hazard functions of the gamma-Weibull model and logistic model.

Word learning as accelerated/decelerated sampling until a given threshold

We assume that word acquisition is a product of experienced events. Formally, then, acquisition can be assumed to be a function of the rate of sampling (from the learning environment) of those relevant events until some threshold for acquisition is passed. By this conceptualization the relevant parameters for the acquisition of any single word are the sampling rate of these learning events and the threshold. From this, we can derive that the duration to reach the threshold for acquisition will follow the gamma distribution. Let f and N be the sampling rate and the threshold until acquisition of word i, and also let M be the cumulative number of total events. Then we obtain the probability that the amount of exposure to the word, k, is larger than N as follows.

$$P(k \geq N) = \sum_{k=N}^{M} \Gamma(M+1)\Gamma(k+1)^{-1}\Gamma(M-k+1)^{-1} f^{k}(1-f)^{M-k}$$
$$= \Gamma(M+1)\Gamma(N)^{-1}\Gamma(M-N+1)^{-1}\int_0^f t^{N-1}(1-t)^{M-N}\,\mathrm{dt}$$

where $\Gamma(N)$ is the gamma function, and the second line is equivalent transformation using the incomplete beta function. When the cumulative number of events M is sufficiently large, the cumulative beta distribution can be approximated by the cumulative gamma distribution. Besides, we assume $f(M\text{-}N)=(\delta^{1}T)^{d}$ that the development of number of events (M-N) follows a polynomial function of time T with a constant δ and the exponent d. Then we obtain an extended form of the cumulative gamma distribution: $P(T;\delta,N,d)=\Gamma(N)^{-1}\int_0^x t^{N-1}\exp(-t)\mathrm{dt}$ where $x=(\delta^{1}T)^{d}$ and δ, N, d>0. The δ is also the sampling rate parameter, N is the threshold of sampled events until acquisition, and d is the exponent of polynomial function of T which indicates the efficiency of the sampling events per unit of time.

Special cases The cumulative distribution $P(T;\delta,N,d)$ is an extension upon both Weibull and gamma distribution. In the case $N=1$, the number of required exposures for acquisition (the threshold) is one, it follows the Weibull distribution. In the case $d=1$, the development of sampled events is at a constant rate, following the gamma distribution. In the case d>1, the number of sampled events as a function of time increases (acceleration), and in the case d<1, it decreases (deceleration). In the case N=d=1, it follows the exponential distribution. Thus, $P(T;\delta,N,d)$ is considered as a hyper model upon the exponential, gamma and Weibull distributions. In sum, the exponential or gamma model indicates learning with a constant sampling rate over time; on the other hand, the Weibull model or .hyper model indicates learning with accelerated or decelerated sampling rate over time.

The logistic model The logistic model is defined as follows: $P(T;\alpha,\beta)=(1+\exp(-\alpha(T-\beta)))^{-1}$ where the α is the sensitivity parameter and β is threshold parameter. Although the logistic model also has a threshold, its threshold β is directly dependent on the independent variable T (e.g., age). In contrast, the threshold in the gamma-Weibull model N is on the sampled number of events which may vary for a given T.

Analysis: Vocabulary Growth Curves

As described earlier, we analyzed the growth of acquisition of individual words among a population of children. Since different kinds of words may show different growth curves that are informative about the underlying processes, we also examine different classes of words.

Method

Acquisition The growth curves were derived from the normative data on productive vocabulary growth from 15 to 30 months of age. These data were collected from parental reports of children's productions and are the normative basis for the MCDI, a parent checklist widely used to measure individual children's vocabulary development (Fenson et al., 1993). The MCDI list includes monthly acquisition rates of

654 words. The words are divided into 21 lexical classes *Action Words, Animals, Body Parts, Clothing, Connecting Words, Descriptive Words, Food and Drink, Furniture and Rooms, Games and Routines, Helping Verbs , Outside Things, People, Places to Go, Pronouns, Quantifiers and Articles, Question Words, Small Household Item, Sound Effects, Toys, Vehicles, and Words about Time*. For this analysis, we use these lexical categories to define different kinds of words. So the dataset to be fit by the models is the proportion of children reported to know each of the 654 words at 15 monthly intervals (16 to 30 month of age).

Properties of words To help us understand the meaning of the fits we also considered properties of words that have been shown in other studies to be related to acquisition, these are adult judgments of AoA, word frequency, familiarity, as given in the MRC database (Coltheart, 1981), frequency of caregivers' speech in the CHILDES corpus (MacWhinney & Snow, 1990), the number of associations for each word in the University of South Florida Free Association Norm (Nelson, McEvoy, & Schreiber, 1998), the number of semantic categories for each word according to Roget's thesaurus (Roget, 1911), the number of synsets (synonym sets) in WordNet (Miller, 1995), and the imageability of each word (Cortese & Fugett, 2004).

Analysis Acquisition rates for each word, defined as the proportion of children who have acquired the word at each month of age from 16 to 30 month olds was fit to each model: hyper gamma-Weibull model, its subsets (i.e., the exponential, Weibull, and gamma model) and the logistic model. The analyses of all words share the common set of independent variable T={16, 17, …, 30} for 16 to 30 month of ages. The hyper model has three parameters, the Weibull, gamma and logistic model have two parameters, and the exponential model has only one parameter for each word.The parameters in the model are estimated by maximizing the likelihood of models to provide the given number of children who have acquired each word at each month. The likelihood is given as follows: $L = n\sum_{i,m}\{p_{im}\log(q_{im}) + (1-p_{im})\log(1-q_{im})\}$ where p_{im} and q_{im} are proportion of children who have acquired the given word i until month of age m in the MCDI and the model respectively. And n=1800 is the number sampled children (Fenson et al., 1993). Since these models have different degree of freedom, we compared their BIC (Schwatz, 1978) calculated as log-likelihood with penalty on the number of parameters: $\mathrm{BIC} = -2 \times L + \log(n_0) \times k$ where k is the number of parameters and n_0=1800*654*15 is degree of freedom of the data.

Results

In order to determine which model is the best descriptor, we compared the degree-of-fit of the five models. The best fitting model is the hyper model ($BIC = 1.7335 \times 10^7$, fitting the best for 45.0% of words). The better models follows in order of the gamma model ($BIC = 1.7344 \times 10^7$, the best for 33.6% of words), the logistic model ($BIC = 1.7364 \times 10^7$, fitting the best for 12.0% of words), the Weibull model ($BIC = 1.7375 \times 10^7$, fitting the best for 9.0% of words) and the exponential model ($BIC = 1.9657 \times 10^7$, fitting the best for 0.003% of words). Since all four models but the logistic model derive from a nested class of models, the logistic model and the hyper-gamma-Weibull family are compared. The logistic fits only 12% of words, in contrast, the gamma-Weibull family fits 88% of words. Remember that the logistic model assumes that each word is learned at a constant rate but has a threshold (age) for acquisition. The fact that the gamma-Weibull models fits more words suggests that the growth pattern of most words depends on the number of sampled events rather than merely age. Further, since the gamma-Weibull model fits most of words, we focus on the specific versions of this model.

We next determined, for each *word*, whether its growth function was better fit by the Weibull or gamma model. We analyzed how much proportion of words in each class fits the Weibull model relative to gamma model. The result shows that more abstract words such as "question words" or "connecting words" tend to be fitted with the Weibull model, in contrast, basic level nouns tend to be fitted with the gamma model (Figure 2). Further analysis shows a positive correlation between average acquisition rates across 16 to 30 month olds and the proportion of words in the word class which fits to the Weibull model against the gamma model (Figure 3; R=-0.614, p<0.05) with the gamma model fitting more early words and the Weibull model better fitting late acquired words. In the hyper model with all three parameters, there is high correlation between the threshold N (gamma) and exponent d (Weibull) parameters across all words. The regression equation $\log(d) = -0.59 \times \log(N) + 0.39$ accounts for 96.2% of variance. This power function shows a sort of tradeoff: the threshold is high if the exponent is low and vice versa. Accordingly, this result suggests that there is continuous transition across all words between vocabulary growth fitting the Weibull distribution and that fitting the gamma. This tells us that there are different growth patterns for different words but it does not tell us why.

The gamma model: threshold based learning A further analysis examined the significant parameters in two models. First we analyzed the sampling rate parameter δ in the gamma model. This parameter indicates how many events are sampled at a given time. This parameter, thus, might be expected to be related to the frequency of words in the learning environment. The correlation between the logarithm sampling rate parameters in the model and the word frequency in CHILDES is -0.252 (p<0.01 and n=609). This correlation is higher than the correlation between average acquisition rates in MCDI and the logarithm of word frequency in CHILDES corpus (R=0.03, p=0.39 and n=609). As Goodman et al. (2008) have shown, across all word classes, frequency is not strongly related to age of acquisition. However, the result shows that the sampling rate parameters which are also estimated from the vocabulary growth curve are more strongly related to frequency than is AoA.

Since the previous analysis suggests that word frequency has a positive correlation *within a word class* (Goodman et al., 2008) and our analysis also shows the gamma and Weibull models fit different classes of words. We analyzed the frequency in the CHILDES corpus for each word class (Figure 3). The analysis shows that only several subsets of nouns and verbs have significant positive correlations in words within its word class. The word classes shown as triangles are significantly correlated with frequency ($p<0.05$, the range from 0.3 to 0.7). These classes are those that are learned earlier and are also the classes that are better fit by the gamma model than the Weibull model. The correlation between proportion of words fitting the gamma model and within-class frequency-parameter correlation is 0.61 ($p<0.05$). In brief, the AoA of words in classes best fit by the gamma model are correlated with frequency and sampling rate and these tend to be early learned words, mostly nouns and verbs.

The Weibull model: acceleration of sampling words

A small set of words are best fit by the Weibull model. Many of these words (but not all) are closed class words (Figure 2): The Weibull distribution is characterized by the exponent parameter of polynomial function of age (i.e., d in the model), and the threshold for learning is one (N=1). The exponent parameter indicates how the number of sampled events develops as a function of age (i.e., in constant, accelerated, or decelerated rate). If the exponent d=1, the number of sampled events increases linearly, but in case of d=2, it increases as a square of age. Roughly, the exponent parameters for words range from 1 to 9, meaning that most words tend to be learned faster by older children.

In order to gain insight as to what this exponent parameter might mean in terms of psychological mechanism, we again looked to the how this parameter correlates with properties of words known to be related to age of acquisition. Of all the variables examined, only imageability was correlated (R=0.564, p<0.01, n=351). Within the Weibull model, the exponent d and AoA have a close relationship, since these parameters are estimated from the AoA pattern. Thus, in order to exclude a spurious correlation between the exponent parameter via AoA, we analyzed the partial correlation which subtracts the effect of AoA. Table 1 shows the correlation and partial correlation among AoA, shape parameters, and imageability of all individual words common between the MCDI and Cortese & Fugett's norm (n=351) The AoA in this analysis is defined the average acquisition rates from 16 to 30 months. The AoA of all individual words has positive correlation to the imageability ratings. This means that words acquired earlier tend to be easy to image. This is natural and consistent with the previous result (Cortese & Fugett, 2004). The important point here is whether the parameters in the Weibull or gamma model are more strongly correlated with imageability *higher* than is AoA. The (partial) correlations of the exponent (Weibull) and threshold (gamma) parameters to the imageability of words are shown in bold letters in Table 1. This result shows that only exponent parameter in the Weibull model is more strongly related to imageability than AoA. In sum, the growth of the words fitting to the Weibull distribution (i.e., closed class words) can be characterized with the exponent parameter which is correlated to imageability.

Table 1: The correlation and partial correlation among AoA, exponent or threshold parameters, Imageability of words.

		AoA-Parameter	Parameter-Image.	Image.-AoA
Weibull	Corr.	-0.860	**0.448**	**-0.564**
	Partial.	-0.822	-0.392	-0.088
Gamma	Corr.	-0.542	**-0.463**	**0.448**
	Partial.	-0.422	-0.294	0.264

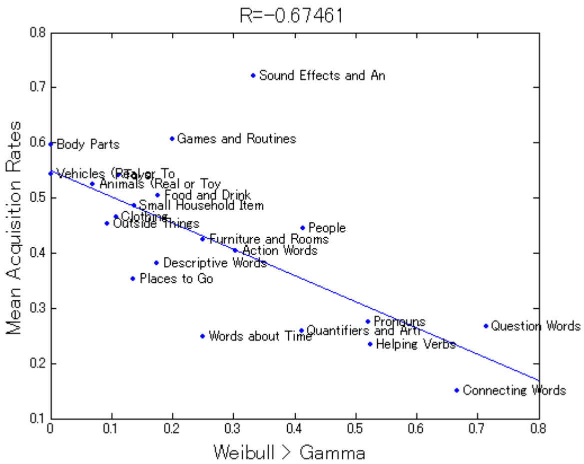

Figure 2: Mean acquisition rates in each word class as function of the proportion of words fitting the Weibull model

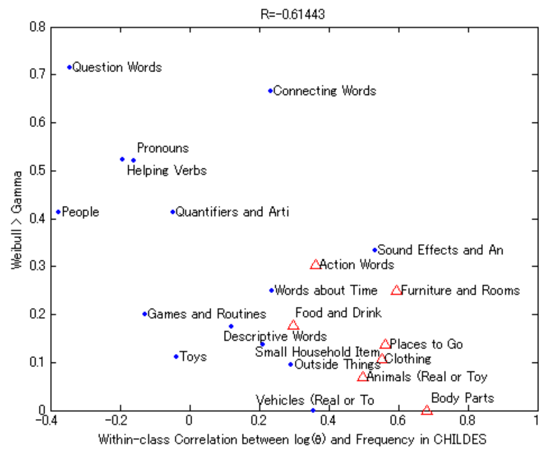

Figure 3: Correlation between the threshold parameter and frequency in the CHILDES in each word class. The red triangle indicates significant correlation.

Discussion

There are three main results. First, the gamma-Weibull class of models fit the vocabulary growth curve better than the logistic model does. This result suggests that, for most words, acquisition depends on the sampling of some events rather than simply on age. Second there is a continuous transition from gamma-type growth in earlier learned words to Weibull-type growth in later acquired ones (Figure 2). Early learned show a constant learning rate, but later learned word classes show an accelerating rate. This suggests that words in different classes or words learned in different periods are learned in different ways. Third, from the detailed analysis of parameters, the two main parameters in the model are correlated with different properties of words, frequency versus imageability.

Finally, we raise a possible underlying learning process of different classes of words from all results mentioned above. For nouns but not function words, the relevant events in the learning environment may be naming events. This is a reasonable interpretation because the sampling rate in the model is highly correlated to frequency in the input for these words. For the other words, which are more of a mixed bag but include function words, the key events may not be the target word alone but also its relation to other words. One possibility is that the exponent parameter may reflect how many words in the learning environment are relevant in this way to learning the target word. This conjecture stems from the present observation that the growth curves of the function words tend to be characterized by the acceleration of sampling; a the number of possible combinations of words would grow faster as a function of words already known (e.g., the number of possible pairs of words grow in quadratic order). This construes the correlation between imageability and exponent parameters. A more relational word with high exponent parameters might be less imageable and vice versa, because such a relational word has fewer meanings itself but has meaning only in relation to other words.

Acknowledgments

This study was supported by grants from NIH MH60200.

References

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of child language* (p. 96-151). Oxford: Basil Blackwell.

Bloom, P. (2000). *How child learn the meanings of words*. Cambridge, MA: The MIT Press.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497-505.

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *36* (3), 384-387.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28* (1), 125-127.

Fenson, L., Dale, P., Reznick, J. S., Bate, E., Hartung, J., Pethick, S., et al. (1993). *MacArthur communicative development inventories*. San Diego: CA: Singular Publishing.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the macarthur communicative development inventories. *Applied Psycholinguistics*, *21* , 95-116.

van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, *105* (4), 634-677.

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, *75* (4), 1098-1114.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35* (1), 515-531.

Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. *Child Development*, *58* , 1523-1531.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language*, *17* (2), 457-472.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317* (5838), 631.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name-nameless category (N3C) principle. *Child Development*, *65* , 1646–1662.

Miller, G. A. (1995). WordNet: An on-line lexical database for English. *Communications of the ACM*, *38* (11), 39-41.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available from http://www.usf.edu/FreeAssociation/

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, *48* , 21-69.

Roget, P. M. (1911). Roget's thesaurus of English words and phrases (1911 ed.). Available from http://www.gutenberg.org/etext/10681

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6* (2), 461-464.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. Psychological Review, 114 (2), 245-272.