Using the words toddlers know now to predict the words they will learn next

Nicole M. Beckage (Nicole.Beckage@Colorado.edu)
Department of Psychology and Neuroscience, 345 UCB
Boulder, CO 80309 USA

Eliana Colunga (Eliana.Colunga@Colorado.edu)
Department of Psychology and Neuroscience, 345 UCB
Boulder, CO 80309 USA

Abstract

We set forth to show that lexical connectivity plays a role in understanding early word learning. By considering words that are learned in temporal proximity to one another to be related, we are able to better predict the words next learned by toddlers. We build conditional probability models based on data from the growing vocabularies of 77 toddlers, followed longitudinally for a year. This type of conditional probability model outperforms the current norms based on baseline probabilities of learning given age alone. This is a first step to capturing the interaction between a child's productive vocabulary and their learning environment in order to understand what words a child might learn next. We also test different types of variants of this conditional probability and find that not only is there information in words that are learned in proximity to one another but that it matters how models integrate this information. The application of this work may provide better cognitive models of acquisition and perhaps allow us to detect children at risk for enduring language difficulties earlier and more accurately.

Keywords: word learning, semantics, language acquisition, co-occurrence, development, longitudinal data, CDI

Introduction

Do children learn words systematically? There is a lot of evidence that words are not all learned equally. Perhaps not surprisingly, for example, parents' vocabulary is related to their children's vocabulary (e.g., Weizman & Snow, 2001; Veen, et al., 2009). That is, the child will learn the words in his or her environment. In addition, some concepts, and therefore the words that name them, may be easier to learn than others. For example, concrete nouns are learned earlier than verbs and adjectives (e.g., Sandhofer, Smith, & Luo, 2000; Gentner, 2006). Furthermore, the child may bring some preferences and constraints to the task of word learning. For example, children may become particularly interested in dinosaurs or construction equipment or even tea sets (DeLoache, Simcock, & Macari, 2007). That is, in characterizing the forces that guide word learning, there is evidence that at least three distinct but not necessarily mutually exclusive sources of information can come to bear: a) the structure and composition of the linguistic environment, b) the structure of the concepts and categories being named, and c) the characteristics of the learner itself. In this paper we focus mostly on this third source of variability by constructing conditional probability models from longitudinal trajectories of word learning that make predictions at the word level, for individual children. That

is, we ask: can we use the words a child knows now to predict the words that a child will learn next?

Measuring the developing lexicon

One well-established way to characterize toddlers' lexicons is to use vocabulary checklists, such as the MacArthur-Bates Communicative Development Inventory (CDI; Fenson et al., 1994). These parent-reported measures have been shown to be effective in evaluating children's communicative skills up to 30 months of age (e.g., Thal et al., 1999). The CDI: Words & Sentences Toddler form is a checklist of over 700 early words that at least 50% of children typically say at 30 months of age. By pooling data over thousands of children, the CDI provides norms of the percentage of children who say each of these words at a given age from 16 to 30 months of age, month by month. Aside from being shown to be a valid measure of communicative skills for this age group, the CDI has been recently shown to be an effective tool for sorting toddlers at the low and high end of the acquisition distribution into late talkers and typically developing children (Heilmann, et. al, 2005). This might allow us to see universality in learning but it also masks some of how the process works—the aggregate cannot explain individual differences but models of learning necessarily must.

The CDI norms can be used to build models of growth. For example, Hills and colleagues (2009) have used CDI norms to build growth models based on networks of words connected by feature similarity or associative strength. Beckage, Smith and Hills (2011) used the connectivity of language within the vocabulary of young learners and showed that there are differences in the structure of the vocabularies of children at risk for language impairments and those of typically developing children.

Note that these approaches presuppose that there is information in the relationships between words. If this is the case, there should be predictive power in looking at the between-word dependencies over time. We do this by exploiting the statistical regularities present in the developing vocabularies of 77 children, followed longitudinally for a year, at monthly intervals.

Rationale

We propose a simple way of uncovering the interaction between the language environment and learning and thus uncovering more of the systematicity of word learning. Instead of just considering the frequency of production for a

given word conditioned on the age of a child, as with the CDI norms, we suggest that there might be additional information in the structure of the language knowledge itself, in the set of words that are known. To pursue this claim, we build the most naïve notion of relatedness and leverage this information in order to predict what words a child might learn next. We define relatedness to be the conditional probability of a word given the child knows another word. For the sake of this paper, we consider words learned in temporal proximity to be related. We build up these values from the longitudinally collected CDIs by considering a connection between words that are learned at the same time (within the same month). We then compute the conditional probability as follows:

$$Pr(a|b) = \frac{P(a \bigcap b)}{Pr(b)} \qquad (1)$$

For example, we compute the probability of knowing "cat" given "dog" by calculating the probability of a child learning "cat" and "dog" in the same month, normalized by the probability of knowing "dog" in the population as a whole. We can then use a variety of methods to combine these conditional probabilities into a single probability of learning word i given that they know a set of words **J**. That is, for each not-yet-known word, we can calculate the probability that the child will learn that word next, given the set of words the child already knows.

In order to combine the conditional probabilities given the set of known words, we need to integrate over the conditional probability given each of the words known. Here we test three different models of this: the **Additive model** assumes that every conditional probability contributes equally. In the additive model we simply sum up the conditional probability of i given every j in the set of known words. This gives us a proportion of learning for every word not yet learned. The issue with this model is that it requires a large amount of information and storage. One rudimentary simplification would be to assume that only the maximum conditional probability was used. This model we call the **Maximal model** because we use only the strongest conditional probability between i and some j from the set of known words **J**. Finally in the **Threshold model** we compare a model that considers conditional probabilities in an additive fashion but considers links only as present or absent. The link is determined as present when the conditional probability strength is above a certain threshold (in our case the median of all conditional probability values) and absent otherwise. We compare these conditional probability models to two population-based models, one based on the CDI norms (**norm-baseline**), another based on the observed frequencies (**observed-baseline**), as well as the **null model** (the assumption that all words will be learned with equal chance). We evaluate the conditional probability models by comparing their predictive power to the population-based models, and use 5-fold cross-validation to evaluate the model's performance in predicting untrained trajectories.

Methods

Vocabularies and Co-occurrences

We utilize CDI measured vocabularies collected at the University of Colorado Boulder. Seventy-seven toddlers between 15.7 and 18.6 months (mean starting age 17 months) were recruited as part of a year-long longitudinal study. These participants completed monthly behavioral tasks as well as vocabulary assessments. The vocabulary assessments were conducted through parent report using the CDI Words & Sentences toddler forms. These CDIs were collected for 12 consecutive months with the majority of parents completing the forms each month. On average we have 9.8 months for each child.

In our study, we include a total of 650 words from the full form, marking duplicate words with parts of speech (such as "orange" as a noun and "orange" as an adjective). We included words that were both on the full form and had norms available online (http://www.sci.sdsu.edu/cdi/). Words that were not part of our modeling included words like *above*, *after*, *on* and *off*. All together we have 77 children and a total of 684 CDI forms. For the sake of this paper we consider each month to be independent of every other month. That is, we build associative structure only from words that are learned during the same month (or words that are known at the beginning of the study.) This limits the co-occurrence measure to capture only short-term dependencies. In the future we plan to extend this work to include cross-sectional vocabularies as well, which will allow us to capture long-term dependencies.

To derive the strength of connectivity, we simply take a count of the number of times two words appear in the same vocabulary (i.e. are learned in the same month) normalized by the population level knowledge as measured from our sample for the words. This provides the basic counts that are then used to compute an 'activation level' that will then give rise to predictions of the next word learned. We then calculate the probability of learning word i given that a child already knows word j. This is then compared to the models based only on population level data as well as a model that assumes uniform learning.

Models

We compute two population-based measures. The first normed model is based on the CDI norms where we consider the likelihood of a child learning the specific set of words we observe to be a function of the population level age of acquisition (AoA) norms (Dale & Fenson 1996). The second measure is calculated analogously, but computing the likelihood according to the AoA as observed in our own sample. We also compare these and all other models to a straw-man baseline measure (the null model) that gives every unlearned word equal probability of being learned.

We compare these population level models to conditional probability models. For the **additive model** we calculate the probability that each word is learned as proportional to the sum of the conditional probability of all known words.

$$Pr(a|\mathbf{B}) \propto \sum_{b_i \in \mathbf{B}} \frac{Pr(a \bigcap b_i)}{Pr(b_i)} \qquad (2)$$

In effect, the probability of learning word *a* is computed such that we sum across the conditional probability of each known word *b* in the set of known words **B**. For example if a child knows words "cat" and "dog" then the probability that the child will learn word "pet" is proportional to the sum of the probability of learning "pet" given "cat" plus the probability of learning "pet" given "dog". This assumes a level of independence that does not exist in language itself.

Using similar methods, in the **maximal model** we consider only the maximum conditional probability to be proportional to the probability of learning a given word. We test this model because it requires only one point of information per word as opposed to considering all possible combinations of known and unknown words.

$$Pr(a|\mathbf{B}) \propto \arg\max_{b_i \in \mathbf{B}} \left\{ \frac{Pr(a \bigcap b_i)}{Pr(b_i)} \right\} \qquad (3)$$

This simplification may still capture much of the variance if the maximal connection dominates the additive model or if strong connections in the learning environment really do highlight words to be learned next.

We also consider thresholded conditional probability. The idea here is that the learner has access to most of the conditional probability space but only at a coarse level. The learner is considered to maintain only the strongest conditional probabilities and that these are considered as present or not. Mathematically, the **threshold model** is:

$$Pr(a|\mathbf{B}) \propto \sum_{b_i \in \mathbf{B}} \mathbb{I}_{\{Pr(a|b_i) > c\}} \qquad (4)$$

Here **I** is the indicator function and is valued only when the conditional probability is greater than some constant c. For this analysis we let c be the median conditional probability across all children. This adds an additional variable to our model but it is set a priori and thus has little significance on the complexity of the model. From an information processing point of view, this model may take less effort since we consider only the presence and absence of a link and not the weight, reducing the complexity of the space.

While we do not consider it here, even these very simple models can quickly be extended to other types of more sophisticated models. First, conditional probability is at best a first order approximation to the full complexity that language embodies. The model can only be as good as the measure used to inform it which in this case is simple co-occurrences. Further, these co-occurrences are based on one month time slots, we could consider data at other time scales, or multiple timescales. Finally, we have chosen to integrate the conditional probability information here in the simplest ways. These models can be seen as network models which allows us to consider not only what words are predicted to be learned next but how these mechanisms for learning transform the semantic structure present in the network. We mention this to highlight the implications of testing these basic assumptions on the larger word learning models (e.g., Vitevitch, 2008; Hills et al, 2009).

Evaluation

We begin by looking at the percent of vocabularies better fit than the null model and the percent likelihood improvement. This tells us some information about the general variability of the input and the ability of the model to account for this variability but little about which model is best. Thus we consider the percent of vocabularies that are better than each of the population based models. This tells us the proportion of vocabularies better fit by the model but not how much better (or worse) of a fit the models give us for our sample. Thus we include the total likelihood of the test data given each of the models. We then compare the likelihood fits across models in order to understand how a model is performing in comparison to other models. We also look at the percent of vocabularies best fit by a given model.

This gives us a good deal of information about the performance of the models and the ability of the models to utilize and combine information in order to predict words learned next at the level of an individual child's vocabulary. To be sure that we are capturing actual signal we want to calculate the conditional probabilities based on a different set of vocabularies than those on which we test the models. Thus, we use cross validation and iteratively build up the necessary associations that the models require from 80% of the vocabularies and then test on the remaining 20%. We do this at the child level since sequential vocabularies are not independent of the child (an issue we've ignored up to this point). We randomly select the 20% test group and repeat this 5 times such that every observed set of vocabularies for a given child is in the test set once. This allows us to test how well the model can predict the vocabulary growth of a child it has no direct information about. We compare the average performance on the five different test sets.

Results

We know what specific words a child learned in a given month, and we use our model to calculate the probability of a given set of words. Some models give a zero probability to learning certain words and thus we first want to look at what percentage of our population cannot be fit by a specified model. This will give us information about how constrained the models are in their ability to fit the wide range of data present in our sample. Column 1 of Table 1 shows the results. In general the models are able to capture the learned words fairly well. The worst model is the Threshold model which is due to the fact that many words are assigned a zero probability under this model since connections that are not above the median strength are considered absent—this results in about 4% of the observed vocabularies not being explainable under the strictest definition of the model. The model based on the CDI norms also has some difficulty accounting for some (2%) of the vocabularies seen. In practice, this means that some children learned words earlier than the normed CDI measures would have predicted—that is, in the vocabularies used to build up the norms, there were no children in the sample that learned some words that children in our study did. This is even more extreme for our

observed CDI norms—which is probably an effect of sample size. In general a large majority of the 684 vocabularies could be fit under the models we constructed; a total of 633 overlapped across all models. We constrain our model comparisons to this subset in all further evaluations.

The next major question is whether or not the constructed models outperform the null model in which each word is given equal probability of being learned. The answer, in short, is that all models perform better than the null model when we consider the total likelihood across all vocabularies. Further, the minimum number of vocabularies better fit by our models than the null model was 82%. This suggests that there is systematicity to the order in which children learn words. In fact we don't only fit the observed data better we get a fairly substantial improvement in overall likelihood when we utilize these models. We have at last 8% improvement and at most 19% improvement.

Table 1: Model performance compared to null model. We consider % of vocabularies not fit, improvement over null and % of vocabularies better fit by a given model.

Model	vocabs not fit (%)	improvement over null (% llk)	vocabs. better than null (%)
Normed	2.37	14.54	81.99
Observed	3.97	**19.04**	**90.52**
Additive	0	18.22	89.10
Maximal	0	8.39	86.41
Threshold	4.05	18.66	82.94

However, showing that words are not acquired randomly does not answer the question of how individual children build up a vocabulary. Returning to the ideas from the introduction, this does not rule out the effect of the structure of the environment. Children learning words proportional to the frequency they encounter them in the environment could explain these results. This would maintain independence between the words a child knows and the words the child is going to learn next. The two baseline models maintain this independence as well: the model based on the normed CDIs and the model fitting to the observed CDIs. In contrast, the other models assume conditional probability plays a role in prediction of vocabulary growth and uses this to link known words to what words will be learned next. Thus, to get at our original question we want to compare these population level models to the other models that require conditional information. We already have a bit of information about the overall model performance when we look at the total likelihood across all vocabularies. We see that we get the largest improvement in likelihood when we utilize the observed CDIs. And we also see that this model gives us the most vocabularies that are fit better than random acquisition.

The gains resulting from using conditional probability are clearer when we consider which vocabularies were best fit rather than looking at the overall likelihood which could be easily inflated by isolated vocabularies that are particularly difficult for a given model to fit. With cross-validation, the threshold model outperforms all others, as shown in Table 2. This improvement is non-trivial as it accounts for the best model in over 50% of vocabularies. This is maintained when we look across children as well—most children are best fit by the threshold model. The observed norm model does provide the best fit for 22% of the data suggesting that there is some predictive powers in the population level rate and time of acquisition. When we look across the population level models we see that over 70% of vocabularies are better fit by a conditional probability model than by a population level normed model. Critically, this suggests that there is some added information in conditional probabilities.

Table 2: Performance with cross-validation. Overall ability to account for the data as well as percent of vocabularies best fit by a given model. For comparison the model performance is directly compared to population models.

Model	% vocabs best fit	% better normed	% better observed
Normed	7.28		30.82
Observed	22.04	69.17	
Additive	10.23	75.20	54.29
Maximal	10.31	25.63	17.67
Threshold	**50.11**	**77.38**	**60.52**

To show the extent of improvement offered by conditional probabilities, we consider the percent of vocabularies better fit by a given model and the CDI data. In Table 2, column 2 and 3, we see that most of the models perform much better than the normed model with roughly 75% of children being better fit under a given model than the published norms and further that many vocabularies are better fit when compared to the norms based on our particular population of toddlers in boulder. This suggests that the norms may be predictive for some children but that in general accounting for the words that are learned previously as well as the relationship of words that are learned together may help us predict what word a child will learn next. Further, the way we combine the type of joint information about word learning may influence our ability to capture vocabulary growth.

Discussion

These results suggest that conditional probabilities do aid in accounting for word learning trajectories. That is, the words that a child already knows can help predict the words that they are going to learn in the future. This implies that there is some sort of systematicity in word learning and that it is not explainable by structure of the environment alone or by conceptual complexity but rather by the interaction of the structure of concepts and meaning within the knowledge of the individual child. The two models that are based on normed data can be seen as independent of individual variation. That is, for these models to perform well at predicting what words a child will learn next, children across a variety of settings and in a variety of learning

environments would be expected to learn words in similar proportion and at a similar rate. This could suggest that the input is structured in a systematic way or that the learning strategy is the same across all children and not dependent on the child's productive vocabulary at any point in time. We did see that these models in general can be fairly predictive of word learning and in fact the total likelihood of the data was minimized under the model that built norms from CDIs collected in our lab. This suggests that these models capture some important aspect of learning. However if we are interested in understanding the different styles of how children learn and capturing the variability across children, these models, inherently, cannot help us with these types of questions as they average out variability.

Looking more closely at the overall likelihood of the models, we see a strong trend that the population models are not as able to adapt to new data. When we fit the models on the full data (that is we included the test set in the training set) the observed CDI norms had a much better total likelihood. However, this model took a big hit in the cross-validation method (results not shown in this paper) suggesting that the observed CDI norms may have overfit the data. The fact that conditional probability models performed better than the population level models in predicting unseen data suggests that the whole story is not in the input alone, but that there is an interaction between a child's productive vocabulary and what words the child will learn next. Even with very simple models of conditional probability we were able to increase our ability to predict and account for the ways vocabularies expand. Thus, if we were to refine our models to include other types of relationships (or more meaningful semantic relationships) between known words and words learned we might be able to understand how children take in their language environment and combine this with their individual vocabulary knowledge to learn new words. The work presented here only begins to look at this by testing models that combine the relationship of co-learned words in different ways, but refinement on these types of models could provide a way for us to uncover not just how children learn new words but also how they integrate a variety of information in order to develop representations of the world. For example, here we considered the median and as our cutoff in the threshold models, but in theory this could be a free parameter fit at the level of individual children (or at the population level conditioned on age) and could hold added information about how children interact with the learning environment. It is true that this threshold model has an additional variable but by setting this before looking at the data we have dealt with any issues in comparing this model to the other models. In the future we plan to do more extensive parameter fits as well as extend the basic models in complexity. For example, we would like to allow the number of maximal values included in our maximal model to be n instead of just 1, where n is a free parameter itself.

The first model (the additive model) tested combined conditional probabilities by maintaining connections and

weights and summing up all of the conditional probabilities between the word candidate and all known words. This resulted in a model that was able to fit much of the data and often better than the population level models. But this was not the best fitting model suggesting that this model might have required too much information, accumulating a ratio that included significant noise in addition to the signal. A huge simplifying assumption that led to our next model was one that suggested that children would maintain only the strongest relationship between a word candidate and known words. This model performed poorly—returning a total likelihood significantly worse than the normed models and the closest to the null model. However, the children's vocabularies that were better fit by this model than the CDI norm models were vocabularies that were often best fit overall by this model. The best model is the model that forces a threshold on the conditional probability matrix. This suggests that strong connections may be the important ones and that the weight of the connection is not important just that it is present.

We do not only gain insight from looking at what models succeed but also what models failed and how. The CDI norming data had difficulty capturing individual vocabularies. It is important to note that in some way this model was handicapped from the beginning. None of the observed data was used in building up the norms. On the other hand, the frequencies noted in the norms were accrued over thousands of children, as opposed to our much smaller sample. Nonetheless, even when the other models were handicapped in the same fashion, the discrepancy in performance still exists. This highlights one of the major weaknesses in utilizing normed data in order to help predict future vocabulary progression. First, it fails to exploit the temporal dependencies available when using longitudinal data. Second, it fails to utilize the dependencies between subsets of words. Of course the poor performance of the norm baseline could be due to a variety of other reasons which would plague any attempt to characterize universals from individual data, and which pose problems to the traditional norming studies. For example, geographic changes between where the norms are collected and Boulder, CO, where our vocabularies were collected could produce variation thus restricting the generalizability of the norms. Or there could be cohort differences due to the fact that the world in which our current children are growing up has a different underlying structure in small but significant ways than the world of the children who contributed data to the norms 20 years ago. This suggests a need for us to consider other tools and methods in order to build up a robust and predictive measure of infant word learning.

Conclusions and Further Directions

Altogether, our findings demonstrate that the conditional probabilities contain information that captures the relationship between the words known by a child at a specific time point and the words that child will learn next. Further, our results show that it matters how we integrate

these probabilities. For example, the maximal model is utilizing only minimal information from the conditional probability (the strongest conditional probability between known and candidate words only) and this model performs very poorly. This suggests that, even though conditional probabilities do contain useful information, not every use of it improves predictive power. The fact that the threshold model does best, suggests that understanding how to combine information can increase fit of the model and allow us to make more accurate future predictions. Interestingly, the model that integrated over the complete conditional probability matrix did not perform better than the model with less information. This result is not atypical for the world of child language acquisition and suggests that perhaps taking into account memory or other cognitive constraints may be useful, if not necessary, in capturing early learning (e.g. Phillips & Pearl, 2012).

This work offers evidence that word learning is affected by a combination of forces and understanding these forces may allow us to predict words that a child would be likely to learn next. We would like to extend these results. Specifically we would like to more closely examine what types of relationships might exist and ways to measure them. If we understand the language environment where a child is learning as well as the way in which the child might be integrating this information with their current vocabulary we should be able to predict which words a child may learn next. This matters because this may allow us to capture children who have learning strategies leading to language difficulty or impairment. These types of models could let us diagnose such children earlier and may allow us to provide effective and child specific interventions.

Another potential direction is the development of tools and techniques that allow us to understand temporal dependencies at different time scales other than a month. Time series analysis combined with graph clustering on the semantics may allow us to expand this work from joint probability to a more complex probability space giving us better temporal resolution as well as more predictive models. Along those same lines, we may be able to fine-tune these models with cognitive theory (which are not included at all in these models, see Hills et al, 2009 for a paper that does consider this) to test generative and process motivated theories of word learning. This would allow us not only to build new computational tools but to refine and expound upon theories of word learning.

At the onset of this paper we asked whether it would be possible to predict the words a child will learn next from the words she knows now. Our findings, even with this simple set of models, suggest that the answer to that question is yes. Significantly, this opens up doors that have far-reaching implications. If we understand how children utilize their environment, conceptual understanding and semantic connectivity as they interact with the world and build up a vocabulary, we can design individualized teaching paradigms that may allow us to build upon, or compliment, what the child already knows aiding in language acquisition.

References

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, *6*(5), e19348.

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

DeLoache, J.S., Simcock, G., & Macari, S. (2007). Planes, Trains, Automobiles —and Tea Sets: Extremely Intense Interests in Very Young Children. *Developmental Psychology, 43(6)*, 1576–1586.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., & Stiles, J. (1994). Variability in early communicative development.*Monographs of the society for research in child development*.

Gentner, D. (2006). *Why verbs are hard to learn*. In K. Hirsh-Pasek, & R. Golinkoff, (Eds.) Action meets word: How children learn verbs (pp. 544– 564). Oxford University Press.

Heilmann, J., Ellis Weismer, S., Evans, J., & Hollar, C. (2005). Utility of the MacArthur–Bates Communicative Inventory in identifying language abilities of late talking and typically developing toddlers. *American Journal of Speech-Language Pathology, 14,* 40–51.

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

Phillips, L., & Pearl, L. (2012). "Less is More" in Bayesian word segmentation: When cognitively plausible learners outperform the ideal. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.

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

Thal, D. J., O'Hanlon, L., Clemmons, M., & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language and Hearing Research*, *42*(2), 482.

van Veen, R., Evers-Vermeul, J., Sanders, T.J.M. & Bergh, H. van den (2009). Parental input and connective acquisition: a growth curve analysis. *First Language, 29(3)*, 267– 289.

Vitevitch, M.S. (2008). What can graph theory tell us about word learning and lexical retrieval? Journal of Speech, Language, and Hearing Research, 51, 408–422

Weizman, Z. O. & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*, 265– 279.