

# Early Talkers and Late Talkers Know Nouns that License Different Word Learning Biases

**Eliana Colunga (eliana.colunga@colorado.edu)**

Department of Psychology and Neuroscience, 345 UCB  
Boulder, CO 80309-0345 USA

**Clare E. Sims (clare.holtpatrick@colorado.edu)**

Department of Psychology and Neuroscience, 345 UCB  
Boulder, CO 80309-0345 USA

## Abstract

In typical development, word learning goes from slow and laborious to fast and seemingly effortless. Typically developing 2-year-olds are so skilled at learning noun categories that they seem to intuit the whole range of things in the category from hearing a single instance named – they are biased learners. This is not the case for children below the 20<sup>th</sup> percentile on productive vocabulary (*late talkers*). This paper looks at the vocabulary composition of age-matched 18-30-month-old late- and early-talking children. The results of Experiment 1 show that late talkers' vocabularies are more variable than early talker's vocabularies. Crucially, Experiment 2 shows that neural networks trained on the vocabularies of individual late talkers learn qualitatively different biases than those trained on early talker vocabularies. These simulations make testable predictions for word learning biases of late- vs. early-talking children. The implications for diagnosis and intervention are discussed.

**Keywords:** Late talkers; early talkers; computational models; neural networks, vocabulary composition.

## Introduction

There is extraordinary variability in the vocabularies of very young children. A two-year-old in the lower 10<sup>th</sup> percentile may produce around 10 words whereas a two-year-old in the top 10<sup>th</sup> percentile will produce well over 300 (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick, & Reilly, 1993). In general, the course of word learning proceeds from slow, effortful learning of nouns and of the range of things that belong in a category, to very rapid learning of object names. Indeed, typically developing 2-year-olds are so skilled at learning new nouns that they seem to intuit the whole range of things in a named category from a single naming experience. This is not necessarily the case for children below the 15<sup>th</sup>-20<sup>th</sup> percentile on productive vocabulary, or *late talkers*. Why do some children learn words quickly and early and others learn words slowly, maybe even showing effects that persist into adolescence? This paper looks at two possible contributing, and interrelated, factors: noun vocabulary composition and word learning biases

The evidence suggests that children become skilled noun learners, at least in part, because they know about the different kinds of properties that are relevant for categorizing different kinds of things. Typically-developing

children show word learning biases that are specific to different kinds of things: they generalize names for solid objects by shape and names for non-solid substances by material (e.g., Jones, Smith & Landau, 1991; Soja, Carey, & Spelke, 1991).

The evidence also suggests that children learn how to learn nouns – and specifically learn how different kinds of properties are relevant for different kinds of things – as a consequence of learning names for things. Each noun the child learns appears to teach the child something general about how to learn new nouns that name things of that same kind, and critically, at the same time, this learned general knowledge constrains and facilitates the types of nouns the child will learn next. To the extent that this interrelation holds true for children in different ends of the language spectrum – late talkers and early talkers – one might be able to leverage this process to predict outcome. The first step, however, is to show that 1) late talkers and early talkers know different sorts of nouns and 2) that these differences in vocabulary structure lead to differences in word learning biases. This paper presents a first look at these questions by examining the noun vocabulary composition of 18-30-month-old late- and early-talking children and showing that neural networks trained on the vocabularies of individual late talkers learn qualitatively different biases than those trained on early talker vocabularies.

## Vocabulary composition and word learning biases

The relationship between vocabulary composition and word learning biases has been typically characterized in one of two ways: abstract knowledge guides, facilitates and indeed allows word learning (e.g., Soja et al, 1991; Gelman & Bloom, 2000) or the words that have been learned give rise to, create, and in fact constitute generalized knowledge about word learning (e.g., Colunga & Smith, 2005, Samuelson, 2008). We would like to bypass the debate on whether word-learning biases are the egg to the vocabulary chicken or the other way around and focus instead on the interrelationship between these two factors.

In the domain of names for objects and substances, and in typical development, vocabulary structure and abstract knowledge in the form of kind-specific generalizations appear to be tightly coupled. First, the tendency to attend to shape in the specific context of naming artifacts emerges as

children learn nouns, becoming particularly robust around the time children have between 50 to 150 nouns in their productive vocabulary (Gershkoff-Stowe & Smith, 2004). Second, the order of development of these word learning biases reflects the statistical structure of early noun vocabularies, (Samuelson & Smith, 1999; Colunga & Smith, 2005). Third, changing 17-month-olds' vocabulary composition by intensively teaching them names for artifacts yields an early bias to generalize names for artifacts by shape *and accelerates learning of object names outside of the lab*, causing a dramatic increase in vocabulary size for children in the experimental training group but not for those in the control groups (Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002). Fourth, computational models trained on the structure of the average 30-month-old vocabulary, show word learning biases like those of young children when processing new objects (Colunga & Smith, 2005), and further the structure of the training set affects subsequent training, facilitating the learning of some sorts of categories but hindering others (Colunga, in prep). Altogether, these results suggest a developmental feedback loop between learning object names, developing biases to attend to the relevant properties for artifacts, and the learning of more object names.

### Late Talkers

Children below the 15<sup>th</sup>-20<sup>th</sup> percentile on normative measures of productive vocabulary size, so-called late talkers, are not a homogenous group in terms of their developmental outcomes: some catch up (Rescorla, 2002), a few will be diagnosed with Specific Language Impairment, and for some the source of the delay may be environmental (Rescorla, Roberts, & Dahlsgaard, 1997). However, Rescorla and colleagues argue against considering late talkers, preschoolers with specific language impairment, and typically developing children as distinct groups, and argue instead for conceptualizing them in terms of a "language endowment spectrum." Importantly, although there is continuity in vocabulary measures at the group level, the outcome for individual children cannot be accurately predicted on the basis of vocabulary production or comprehension (Thal, Bates, Goodman, & Jahn-Samillo, 1997; Desmarais, Meyer, Bairati & Rouleau, 2008).

The literature briefly reviewed above suggests that, in typical development, the words a child knows and what the child knows about learning words in general go hand in hand, and that learning names for categories of things organized by shape speeds up learning nouns. However, this may not be the case for all children. Unlike typically developing children, late talkers do not systematically extend the name of a novel solid object to other objects that match it in shape, and in fact, in one study, almost half of the late talkers systematically extended the novel name of a solid object to others matching in texture rather than shape (Jones, 2003). The decoupling of vocabulary acquisition and word-learning biases may mean that these children are not just limited in their production of object names (the measure

that defines them as late talkers) but also deficient in the processes that subserve the acquisition of new words and in their knowledge about those categories. If this were the case, a natural prediction would be that noun vocabularies of late-talkers should have a different structure than noun vocabularies of typically developing children. For the purposes of this paper we will focus on contrasting the vocabularies of children on the two opposite ends of the spectrum, late talkers and early talkers.

## Experiment 1

### Method

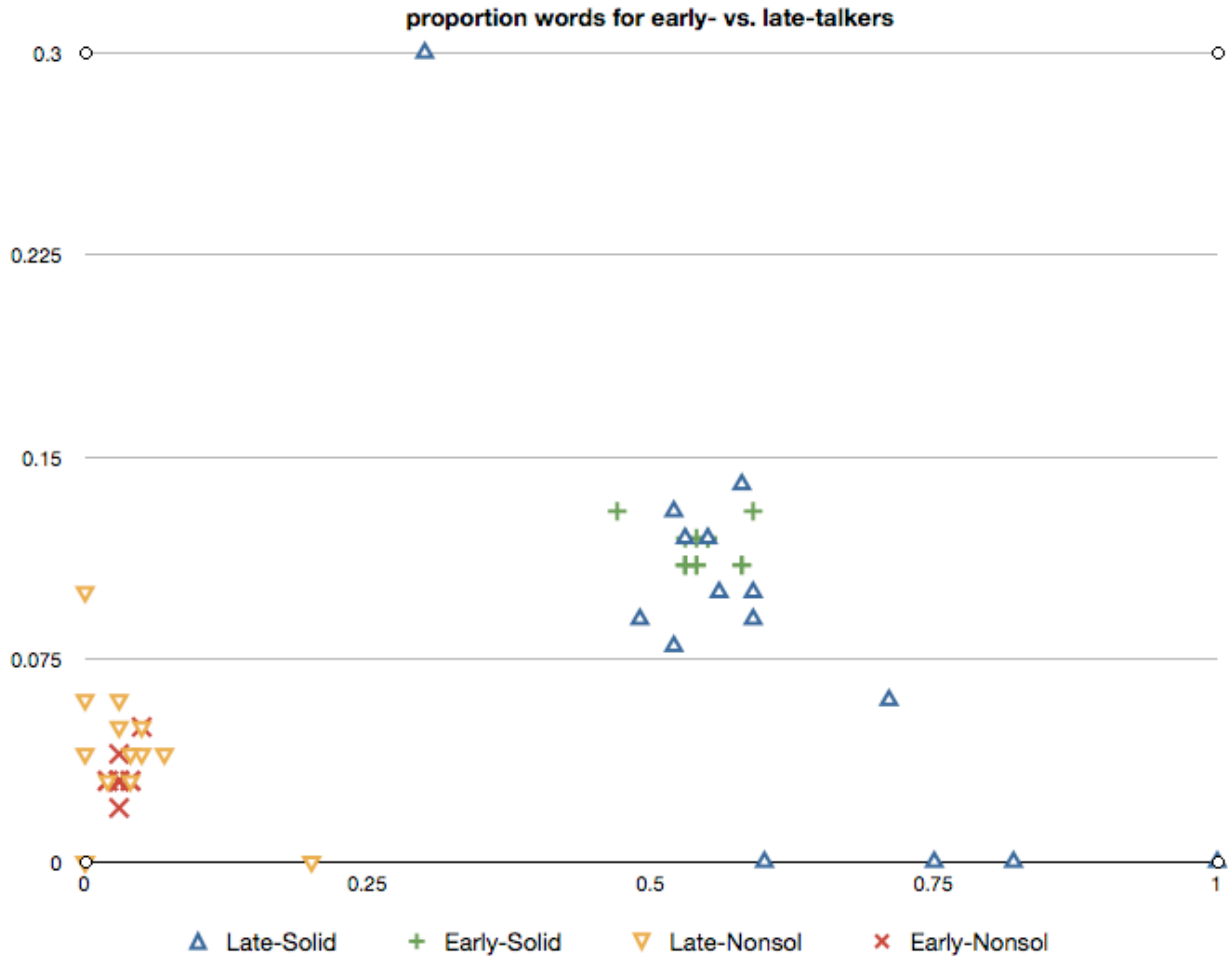
**Materials.** The vocabulary measure used is the Bates-MacArthur Communicative Developmental Inventory toddler version (MCDI) both to select children and to measure vocabulary composition. This is a parent checklist that asks parents to indicate the words that their child *produces* and although it is imperfect as a measurement instrument (Fenson, et al, 1994) it appears to be reliable and to be systematically related to children's performances in a variety laboratory measures of word learning, including especially their word-learning biases in the Novel Noun Generalization task (e.g., Landau, et al, 1988).

**Participants.** The vocabularies of 15 late talkers and of 15 early talkers were selected out of a pool of 148 parent-filled MCDI forms for children between 18-30 months of age. The criterion for inclusion was that there existed a vocabulary form from a child matching in age to within 5 days in both the late talker and the early talker groups. Late talkers fell under the 25<sup>th</sup> percentile; early talkers were above the 75<sup>th</sup> percentile according to the MDCI norms.

The ages for the two language groups ranged from 18.49 months to 28.26 months ( $M=23.14$  and  $23.15$  for late and early talkers respectively). Vocabulary sizes for the late talker group ranged between 15 and 425 words ( $M=132.53$ ); for the early talker group vocabulary size was between 158 and 664 words ( $M=457$ ).

### Results

To get a sense of the variability in vocabulary composition (as opposed to vocabulary size) in children at different percentiles in vocabulary development, for each individual child, we looked at the proportion of nouns they knew for the categories of 1) solid things alike in shape (e.g., spoon), 2) solid things alike in material (e.g., chalk), 3) solid things alike in both shape and material (e.g., penny), 4) non-solid things alike in shape (e.g., bubble), 5) non-solid things alike in material (e.g., milk), 6) non-solid things alike in both (e.g., jeans). Nouns in children's vocabularies were classified as falling in each of these categories according to adult judgments made for each of the nouns in the MCDI reported in Samuelson & Smith, 1999. Figure 1 shows the *proportion* of words for solids and non-solids that are organized by shape (x-axis) and material (y-axis), for each of the 15 late talkers (triangles) and early talkers (crosses).



Although there is some overlap between the two groups, there is greater variability in the composition of the vocabularies of the late-talker sample, for both solids and non-solids. Subjecting children’s proportion of words to a 2(Language Group: early talkers, late talkers) x 2(Solidity: solid, non-solid) x 3(Dimension: shape, material, both) repeated measures ANOVA with age in months as a covariate, yielded the expected main effects of solidity,  $F(1,27)=50.7$ ,  $p<.0001$ , and dimension,  $F(2,54)=8.416$ ,  $p=.001$ , indicating that there were more words for solids than non-solids, and more words for shape-based categories than any other type respectively. In addition, the expected interaction between solidity and dimension was significant,  $F(2,54)=12.37$ ,  $p<.0001$ . There were more shape-based words for solids, and less shape-based words for non-solids. There was also a marginally significant 3-way interaction between solidity, dimension and language group,  $F(2,54)=3.18$ ,  $p=.055$ . Descriptively, late talkers have relatively more words for solid that are organized by shape than early talkers, and relatively fewer words for solids organized by material or both.

### Discussion

As predicted, late talkers and early talkers show a difference in the structure of their noun vocabularies. As a group, late talkers show more variability in their vocabulary structures than early talkers. This is perhaps not strange given that, on average, the children in the late talker group have smaller

vocabularies and thus many more ways of “selecting” the words they know out of the vocabulary checklist. Put another way, as early talkers approximate mastery of the whole checklist, their vocabularies will tend toward the structure of the checklist. The crucial question, then, is whether these differences in vocabulary composition are differences that matter. Do the different nouns late- and early-talkers know yield different word learning biases?

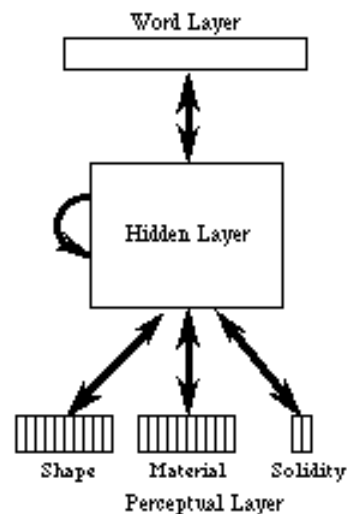


Figure 2. Architecture of the networks used in Exp. 2

To answer this question, in Experiment 2 we trained individual neural networks on the noun vocabulary structure of each individual late-talking and early-talking child in Experiment 1. If the differences in vocabulary structure can, to some extent, explain the differences in language ability, we would expect late talker vocabularies to yield different word learning biases than early talker vocabularies. More specifically, we would expect early talker vocabularies to yield word learning biases that would facilitate the learning of a vocabulary structured like the MCDI – highlighting shape similarities for solids and material similarities for non-solids. In contrast, we would expect networks trained on late talkers' vocabularies to generalize more variable word learning biases, and perhaps even biases that would be unhelpful in learning early vocabularies.

## Experiment 2

### Method

The computational models are a modified version of the ones Colunga & Smith, 2005. The main difference is that these networks were trained using the Leabra algorithm, an algorithm that combines Hebbian and error driven learning (O'Reilly, 1996), instead of Contrastive Hebbian Learning as in the original simulations.

**Architecture.** The architecture is implemented as shown in Figure 2. Words are represented discretely (as single units) and are input on the Word Layer. Referents are represented as distributed patterns over several dimensions on the Perception Layer. For example, the shape and material of an object (say the roundness of a particular ball and its yellow rubbery material) are represented by an activation pattern along the Perception layer. Solidity and Non-solidity are represented discretely; one unit stands for Solid and another for Non-Solid. Finally, there is a hidden layer that is connected to all the other layers and to itself. These networks have been shown to model performance in an analog of the NNG Task when trained on vocabularies structured as those of the average 30-month-old.

**Training.** The networks are trained with categories presenting the same correlational structure as each individual child's noun vocabulary. On each training trial, a word is paired with a referent. The patterns associated with each word are determined by adult judgments of the early noun corpus. For example, adults judged balls to be similar in shape but different in material. To simulate this, we randomly selected an input vector to represent ball shape. Then on individual training trials, we paired that pattern with the label ball and a randomly selected material pattern. We do this for each noun in the training set. Each network was trained in this way for its simulated vocabulary until they reached asymptotic (and near perfect) performance. This part of the simulation is intended to put into the networks the lexical knowledge that the individual child would bring to the laboratory NNG task.

Because we are interested in the consequences of different vocabulary structures regardless of their size, all networks

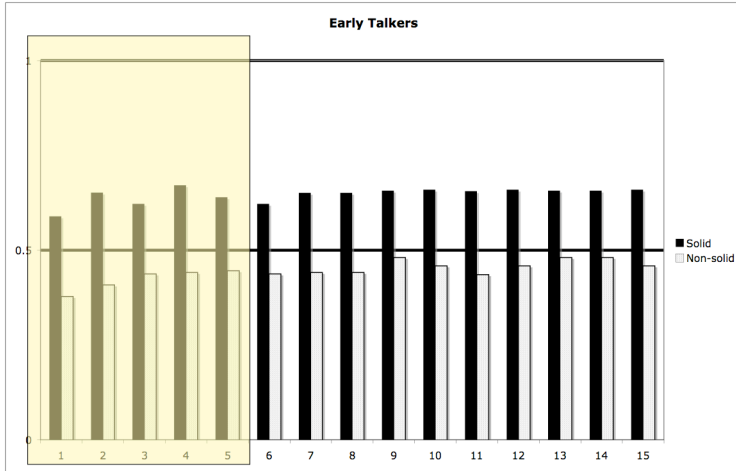
were trained to learn 24 nouns, proportionally structured like their corresponding child's vocabulary. Thus, the only difference between networks were the differences in vocabulary composition found in Experiment 1.

**Testing.** The question is what sort of word learning bias will the networks learn given different vocabulary structures. We address this question in a virtual version of the NNG task. On each test trial of the virtual NNG task, we presented the network with three novel entities (one at a time) on the perception layer – an exemplar, and two choice items, one matching the exemplar in shape only and one matching in material only. For each of these three inputs, we recorded the resulting pattern of activation on the hidden layer. This is a measure of how the network represents these items. If the network emphasizes the shape of the item then the similarities of the internal representations for the exemplar and its shape matching choice should be *greater* than the similarity of the internal representations for the exemplar and the material matching choice. If, however, the internal representations highlight the material of the items, then the similarity of the internal representations for the exemplar and the shape matching choice should be *less* than the corresponding similarity of the exemplar and the material matching choice. We used these similarities along with Luce's choice rule to calculate probability of choice using these similarity measures in order to predict performance in the novel noun generalization task.

In previous work these models have been used to demonstrate the plausibility of the idea that the correlations in the early noun lexicon are sufficient to create second order generalizations – knowledge that any solid thing should be named by shape, and any non-solid thing should be named by material. The present simulations extend this work to variable vocabularies of individual children in the bottom and top ends of the language endowment spectrum.

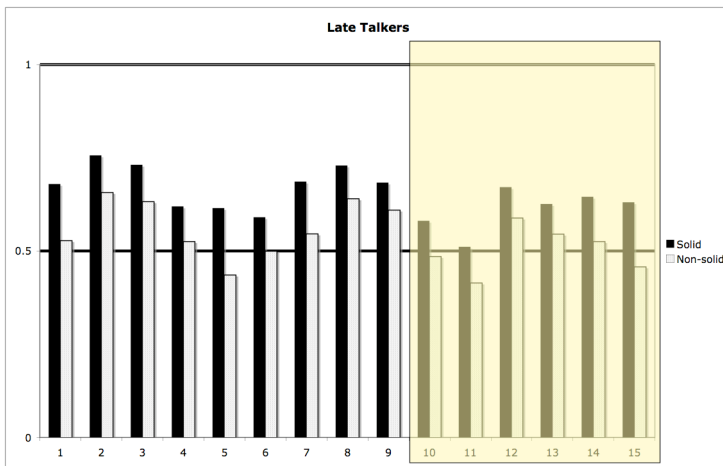
### Results

The networks' predictions for each of the fifteen vocabularies of early talkers and late talkers are shown in Figures 3 and 4 respectively. In short, all networks in the early talker group show a shape bias for solids, and 12/15 early talker networks show a material bias for non-solids as well. In contrast, 12/15 late talker networks show a shape bias for solids and only 3/15 show a material bias for non-solids. Interestingly, 6/15 late-talker networks show a shape bias for *non-solids*, a novel prediction that has not been empirically tested so far. To further analyze the networks' performance, networks were classified according to the observed generalization patterns: *correct* if they showed a shape bias for solids and a material bias for non-solids, *half-right* if they show the appropriate shape bias for solids but no consistent bias for material, or *wrong*, if they showed an incorrect overgeneralized shape biased to non-solids. A chi-square test showed these types of word learning biases were distributed differently in late talker and early talker networks,  $X^2(2,15)=11.21, p=.003$  (Yates'  $p=0.017$ ).



**Figure 3. Predicted proportion of shape choices for each of the early talker networks**

Note that our small sample did not allow for precision matching by vocabulary, however, the highlighted portions of the two graphs above indicate the area of overlap in vocabulary range. The late talker children in the highlighted area have vocabulary sizes between 155-425 ( $M=257.2$ ) and ages between 23-28 months ( $M=25.2$ ); the early talker children in the highlighted area have vocabulary sizes between 158-451 ( $M=331.3$ ) and ages between 18-22 months ( $M=19.9$ ). As shown in the figures, the predicted word learning generalizations for these vocabulary-matched children are qualitatively different – the networks predict that older late-talker children with similar vocabulary sizes as their younger early-talker counterparts nevertheless show less robust word learning generalizations in the NNG task.



**Figure 4. Predicted proportion of shape choices for solids and nonsolids for the late talker networks**

### Discussion.

The results of the simulations suggest that the differences in noun vocabulary composition between late- and early-talking children may result in differences in word learning biases. It is important to note a couple of things that may seem to contradict previous findings. First, the children

whose vocabularies went into these simulations were between 18 and 28 months of age. In general, children at this age do not show a robust material bias for non-solids; that does not happen until age 3 (but see Colunga & Smith, 2005 for an early material bias for non-solids presented in simple shapes in children in this age group), yet a majority of the early talker networks (and a couple of late talker networks) show material biases for non-solids. This is a novel prediction. What the networks suggest is that early talkers will show an early material bias for non-solids. We are currently running a longitudinal study examining vocabulary composition via the MCDI and word learning biases in the lab in children starting at 16 months, and by 19 months of age, every child in the early talking group (6 children) show a robust shape bias for solids and an equally robust material bias for non-solids in a novel noun generalization task. Although this is a small sample size, and there is some question as to whether the effect will remain stable, this preliminary data suggests that the networks' predictions may indeed be true.

The second result from the networks that appears to contradict findings documented in the literature is that a majority of late talkers show a shape bias for solids. This goes against the findings reviewed in the introduction in which late talkers did not show a shape bias for solids and might have even have a texture bias instead. One possible way to reconcile this contradiction is that the children in the Jones, 2003 study were generally older than the children in this study (25-41 months of age,  $M=33.25$ ). It is possible that the shape bias for solids predicted by these simulations will disappear in the next 10 months or so. There is another intriguing possibility, however, and one that points out a limitation in the models. These models do not make a distinction between naming and non-naming contexts. It is possible that the shape preference for solids here is not a true shape bias, but rather an overgeneralized heightened attention to shape. The fact that 6/15 late talker vocabularies yielded a shape bias for non-solids suggests that this may be the case. Looking at the late-talking group of the ongoing longitudinal study suggests that this may be the case, 5/8 toddlers at 19 months show an overgeneralized shape bias for non-solids and for a non-naming control task. A study with larger numbers of early takers and late talkers is necessary to confirm these predictions.

### General Discussion

The work presented here makes several contributions. First, the findings of these two studies show that late talkers and early talkers know different sorts of nouns, a new finding that may have important implications for early identification of at-risk children. At the very least, the finding that there are different vocabulary structures in these two groups of children is a promising direction in looking for ways to predict outcome from characteristics that are easily measured at an early age.

Furthermore, the finding that these differences in vocabulary composition lead to qualitatively different word

learning biases in a computational model that has been previously shown to capture various aspects of novel noun learning, suggests a promising use for process-level computational models. Efforts to tease apart the contributions of different factors to outcomes in late talkers have come up with some characteristics that put children at higher risk, but the underlying mechanisms are not well understood and the need to identify subgroups within late-talking toddlers remains. The work of Ziegler and colleagues in the domain of dyslexia offers a good example of the potential for using computational models – and specifically models that operate at the mechanistic level – in simulating individual differences and further understanding subtypes in atypical development (Ziegler, Castel, Pech-Georgel, George, Aario, & Perry, 2008). Thus, the models presented here are a promising first step in leveraging computational models to aid in the understanding of why some late talkers catch up and others do not.

Finally, these models represent an important extension over previous word-learning modeling efforts in that they go beyond modeling the performance of the mythical average child to making predictions about the performance of individual children, and of children who are both at the top and at the bottom of the vocabulary spectrum. In so doing, the simulations presented here make novel and testable predictions. They predict that early talkers and late talkers will show different word learning biases in the novel noun generalization task. More specifically, the simulations predict that, between 18 and 30 months of age, early talkers will show an early material bias and that late talkers will show an overgeneralized shape bias.

The work presented here also has some clear limitations. First, the fact that we do not have outcome data for the children in these studies seriously constrains what we can infer from these results and their potential use in early identification of at risk children – will the late talking children who show correct biases catch up? Or are the ones showing the overgeneralized shape bias the ones on the right track? Are these differences in vocabulary and in word learning biases predictive of outcome? Second, all of these networks are identical except for the vocabulary structure on which they are trained. Although it is possible to see this as a strong demonstration of the relationship between vocabulary composition and word learning biases, allowing for pre-existing individual differences in these models may increase their power. Finally, there is more to language, and even more to word learning, than learning nouns. Thus, these models capture only a sliver of language learning and may miss components crucial to achieving the ultimate goal of increasing diagnostic power at the individual level.

In spite of these limitations, the models presented here constitute an innovative approach to predicting and characterizing typical and atypical vocabulary acquisition in young children. The relationship between vocabulary composition and word learning biases modeled here – the words you know determine the way you learn new words, which constrains and facilitates the words you will know

next, and so on – opens a new way of thinking about computational models, to capture not only averages and not only individuals, but individual *trajectories*. If we can build computational models that can successfully capture this self-constructing developmental loop, the implications for early diagnosis, designing early interventions, and understanding the mechanisms that underlie word learning in typical and atypical development are far-reaching.

## References

- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: a role for associative learning. *Psychological Review*, *112*(2), 347–382.
- Desmarais, C. S., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *Intl. J. of Language and Communication Disorders*, *43*(4) pp:361-389.
- Fenson, L., Dale, P., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA:
- Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, *76*(2), 91-103.
- Gershkoff-Stowe, L., & Smith, L.B.(2004). Shape and the first hundred nouns. *Child development* *75*(4),1098-1114.
- Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Dev. Science*, *6*(5), 477-483.
- Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*(3), 499-516.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299-321.
- Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research*, *45*, 360-371.
- Rescorla, L., Roberts, J. & Dahlsgaard, K. (1997). Late talkers at 2: Outcome at age 3. *Journal of Speech & Hearing Research*, *40*, 556-566.
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*(1), 1-33.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. . (2002). Creating a shape bias creates rapid word learners. *Psychological Science*, *13*, 13-19.
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning. *Cognition*, *38*(2), 179-211.
- Thal, D. J., Bates, E., Goodman, J., & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late- and early-talking toddlers. *Developmental Neuropsychology*. *13*(3), 239-273.
- Ziegler, J., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition*, *107*, 151–178.