Interaction of Word Learning and Semantic Category Formation in Late Talking

Aida Nematzadeh, **Afsaneh Fazly**, and **Suzanne Stevenson**
Department of Computer Science
University of Toronto
{aida,afsaneh,suzanne}@cs.toronto.edu

Abstract

Late talkers (LTs) — children who show a marked delay in vocabulary learning — have also been shown to differ from normally-developing (ND) children with respect to the semantic organization of their learned vocabulary. We use a computational model of word learning to study how individual differences between LTs and NDs give rise to differences in abstract knowledge of categories emerging from learned words, and how this affects their subsequent word learning. Our results suggest that the vocabulary composition of LTs and NDs differ at least partially due to a deficit in the attentional abilities of LTs, which also results in the learning of weaker abstract knowledge of semantic categories of words.

Introduction

Late talkers (LTs) are children with a marked delay in word learning at an early age, some of whom go on to exhibit specific language impairment (SLI). Early identification of LTs at risk for SLI is especially important, since early intervention can produce significant changes in the language development of these children (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008). Many psycholinguistic studies have thus focused on understanding signs of late talking, as well as factors contributing to it (Paul & Elwood, 1991; Thal, Bates, Goodman, & Jahn-Samilo, 1997; Rescorla & Merrin, 1998; Ellis Weismer & Evans, 2002; Rowe, 2008; Stokes & Klee, 2009).

An important observation about late-talking children is that they seem to not only learn *more slowly* than their normally-developing (ND) peers, but also to learn *differently*. For example, the vocabulary composition of LTs shows greater variability, e.g., in terms of how consistently certain properties, such as shape, are associated with particular categories, such as solid objects (Jones & Smith, 2005; Colunga & Sims, 2011). More generally, the vocabulary of LTs has been shown to exhibit less semantic connectivity than that of NDs (Sheng & McGregor, 2010; Beckage, Smith, & Hills, 2010). The greater variability and the weaker connectivity in the vocabulary of LTs call for further investigation since they might be reflective of underlying cognitive deficits in these children.

Psycholinguistic evidence suggests that children's word learning improves when they form some abstract knowledge about what kinds of semantic properties are relevant to what kinds of categories (Jones, Smith, & Landau, 1991; Colunga & Smith, 2005; Colunga & Sims, 2011). This abstract knowledge is argued to emerge by generalizing over the learned words. Stated otherwise, words that have been learned contribute to generalized abstract knowledge about word meanings and semantic categories, which then guide subsequent word learning. It is possible that because of the differences in the vocabulary composition of LTs and NDs, the two groups of children also form different abstract knowledge of categories, which causes differences in their word learning (as suggested by Jones & Smith, 2005; Colunga & Sims, 2011).

We investigate this possibility by examining within a computational model the precise interaction between early word learning and knowledge of semantic categories of words. We do so by extending an existing model of cross-situational word learning (Fazly, Alishahi, & Stevenson, 2010). As in Nematzadeh, Fazly, and Stevenson (2011), we simulate the difference between ND and LT learners as a difference in the ability of the cross-situational learning mechanism to attend to appropriate semantic features for a word. Within this framework, we propose a new model that forms clusters of words according to their learned semantic properties, and that uses this knowledge in guiding the future associations between words and meanings. We show that the semantic clusters of words are qualitatively very different for our ND and LT models; moreover, the two learners exhibit striking differences in terms of the usefulness of their learned clusters for subsequent word learning. Through computational modeling, we thus suggest an interaction between the impaired ability of LTs to form informative abstract semantic categories, and the observed delay in their vocabulary acquisition.

The Computational Model

Overview of the Word Learning Model

The model of Fazly et al. (2010) is a cross-situational learner that incrementally forms probabilistic associations between words and their semantic properties. The input to a child is simulated as a sequence of utterances (a set of words), each paired with a scene representation (a set of semantic features, representing what is perceived when the words are heard):

Utterance: { *she*, *drinks*, *milk* }
Scene: { ANIMATE, PERSON, FEMALE, CONSUME, DRINK, SUBSTANCE, FOOD, DAIRY-PRODUCT }

Given such an input pair, the model adjusts its probabilistic representation of the meaning of each word. First, the model determines, *based on its current probabilistic knowledge of word–meaning associations*, which semantic features in the scene are more and less likely to be associated with each word in the utterance. Using that assessment of word–feature alignment in the current input, the model then updates its probabilistic representation of the meaning of each word.

In this way, the model uses cross-situational evidence to gradually improve its representation of the meaning of each word w as a probability distribution, $p(.|w)$, over all semantic features: i.e., $p(f|w)$ is the probability of feature f being part

of the meaning of word w. At the heart of this process are two calculations which we briefly summarize here (see Fazly et al., 2010, for more detail). The *alignment probability* determines how strongly a word w and a feature f are associated in the current (multi-word) utterance U at time t, in proportion to the model's current hypothesis of how likely the feature is part of the meaning of the word:

$$a_w^{(t)}(w|f) = \frac{p^{(t-1)}(f|w)}{\sum_{w' \in \mathrm{U}_t} p^{(t-1)}(f|w')} \quad (1)$$

In order to collect this knowledge across all cross-situational uses of the word and feature, the model maintains an incrementally accumulated sum of these alignments that captures the overall strength of the association between w and f:

$$\mathrm{assoc}^{(t)}(w, f) = \mathrm{assoc}^{(t-1)}(w, f) + a_w^{(t)}(w|f) \quad (2)$$

The second key formula to the operation of the model is the *meaning probability* that uses the association scores to update the meaning of each word after processing an input pair:

$$p^{(t)}(f|w) = \frac{\mathrm{assoc}^{(t)}(f, w) + \lambda(t)}{\sum_{f' \in \mathcal{M}} \mathrm{assoc}^{(t)}(f', w) + \beta \cdot \lambda(t)} \quad (3)$$

where β is the number of expected distinct features, $\mathcal{M}$ is the subset of those features that have been observed, and $\lambda(t)$ is a smoothing function which we formulate in a way that captures the developing ability of the model to attend to input, as follows.

Research has shown that children's ability to attend to relevant features of a perceived scene improve over time (e.g., Mundy et al., 2007). Moreover, LTs have been observed to show difficulty with the communicative abilities that enable children to direct appropriate attention on relevant aspects of a scene (e.g., Rescorla & Merrin, 1998). In recent work (Nematzadeh et al., 2011), we demonstrated that we can use the $\lambda(t)$ function to simulate how quickly or slowly the attentional abilities of a learner develop over time. Specifically, the $\lambda(t)$ function determines how much weight is given to unobserved word–feature pairs, with greater weight reflecting immature attentional skills in which the learner fails to focus on the observed (appropriate) meaning features. In the model, $\lambda(t)$ is designed to decrease over time, to simulate gradually improving attentional processes that can appropriately focus on the observed word–feature pairs. We modeled the difference between ND and LT learners by having a $\lambda(t)$ function for the latter that decreases much more slowly, corresponding to delayed development of appropriate attention to the input. Here we adopt that same formulation,[1] but extend the model as follows to consider the role of attention and its interaction with semantic category formation in word learning.

[1]Our ND and LT simulations here use the same settings for $\lambda(t)$ as what we referred to as ND and LT$_{.5}$ in our previous work.

Learning Semantic Categories of Words

We extend the word learning model above by incorporating the ability to form clusters of words based on their learned semantics, and to use the resulting semantic categories in subsequent word learning. [2] These abilities represent a first step in integrating the model's word learning with formation of conceptual categories. These extensions to the model are key to further examination of the cognitive mechanisms that might underlie the weaker semantic connectivity observed in the vocabulary of LTs. Specifically, while Nematzadeh et al. (2011) showed that learned words of their ND learner had greater semantic coherence than those in the LT learner, the model did not actually form semantic clusters of words, nor use semantic relations among words to help in word learning.

Our new model, at certain points in time (depending on the simulation), groups the words it has observed into clusters based on the similarity among their learned meanings. Given two words w and w', we determine their degree of semantic similarity by treating their learned probability distributions over the semantic features, $p(.|w)$ and $p(.|w')$, as input vectors to the cosine function. These cosine values guide the grouping of words using a standard unsupervised hierarchical clustering method. The clusters of semantically related words can then be analyzed to see how the factors that simulate ND and LT learners in the model contribute to different quality of semantic categorization, as observed by Sheng and McGregor (2010) and Beckage et al. (2010), among others.

Moreover, the semantic clusters enable us to build further on the explanation of late talking as arising from attentional differences in learners (as proposed in Nematzadeh et al., 2011). Specifically, we assume that learned semantic categories enable children to generalize their knowledge of related words, which can help focus subsequent word learning on relevant semantic features in the input. In our model, knowledge about the semantic category of a word can be used as an additional source of information about which semantic features are more likely to be aligned with the word in a given input. For example, features such as EDIBLE and FOOD should be more strongly aligned to a word referring to a kind of fruit than to a word referring to a kind of vehicle.

We achieve this in our model by aligning a word w and a feature f in an input utterance–scene pair according to both word-level and category-level information, the latter drawing on the incrementally created semantic clusters. We adopt the formulation used by Alishahi and Fazly (2010) to combine word and category information in the alignment probability: [3]

$$a^{(t)}(w|f) = \omega \cdot a_w^{(t)}(w|f) + (1 - \omega) \cdot a_c^{(t)}(w|f) \quad (4)$$

[2]We continue to refer to the clusters that our model learns both as *clusters*, to emphasize that they are learned in an unsupervised manner, and as *semantic categories*, to emphasize their connection to children's knowledge of abstract categories.

[3]The approach of Alishahi and Fazly (2010) differs from ours: (1) They examine the role of syntactic categories (*e.g.*, noun or verb) in word learning while we look at semantic categories. (2) They use predefined correct assignments of words to such parts of speech, but our clustering is based on the model's learned semantic knowledge.

apple: { FOOD:1, SOLID:.72, ⋯, PLANT-PART:.22, PHYSICAL-ENTITY:.17, WHOLE:.06, ⋯ }

Figure 1: Sample true meaning features & their scores for *apple*.

The first component of the above formula, $a_w^{(t)}(w|f)$ is the word-based alignment, given in Eqn. (1) above. The second component, $a_c^{(t)}(w|f)$, is an analogous category-based alignment (described below). The ω term is a weight (between 0 and 1) that determines the relative contribution of the two alignments; here we use a balanced weighting of 0.5.

Where the word-based alignment captures the association between a feature and a single word, the category-based alignment, $a_c^{(t)}(w|f)$, assesses the overall association between the feature f and the words in cluster(w), the cluster assignment determined by the model for word w. This alignment is calculated by replacing occurrences of $p(f|w)$ in Eqn. (1) with $p(f|\text{cluster}(w))$. We again follow Alishahi and Fazly (2010) in defining $p(f|\text{cluster}(w))$ as the average of the meaning probabilities of the words in the cluster:

$$p^{(t)}(f|\text{cluster}(w)) = \frac{1}{|\text{cluster}(w)|} \sum_{w \in \text{cluster}(w)} p^{(t)}(f|w) \quad (5)$$

where $|\text{cluster}(w)|$ is the number of words in the cluster.

Semantic Representation in the Model

The Representation of a Scene

The input data for our model consists of a set of utterances paired with their scene representations. As in Nematzadeh et al. (2011), the utterances are bags of lemmatized words, taken from the child-directed speech (CDS) portion of the Manchester corpus (Theakston et al., 2001, from CHILDES MacWhinney, 2000). The corpus is transcripts of conversations with 12 British children, ages 1;8 to 3;0. We use half the data as the development set, and the rest for final evaluations.

The corresponding scene representation for each utterance must be artificially generated, since no semantic annotation of the contextual scene exists for any large corpus of CDS. First, we create an input-generation lexicon containing the "true" meaning $t(w)$ for each word w in our corpus: $t(w)$ is a vector over a set of semantic features, each associated with a score. An example lexical entry is given in Figure 1; the creation of this lexicon is described below.[4] Next, to generate the scene S for an utterance U, we probabilistically sample an observed subset of features from the full set of features in $t(w)$ for each word $w \in$ U. This imperfect sampling allows us to simulate the noise and uncertainty in the input, as well as the uncertainty of a child in determining the relevant meaning elements in a scene. The scene S is the union of all the features sampled for all the words in the utterance.

The Representation of Word Meaning

We focus on the semantics of nouns, since they are central to work on the role of category knowledge in word learning. Our previous work (Nematzadeh et al., 2011) used a psycholinguistically-plausible set of features for this purpose (Howell et al., 2005); however, they were only available for a limited number of nouns. Here we develop an improved semantic representation for nouns that enables a more extensive test of our clustering method and associated processing involving semantic relatedness among words.

We construct the lexical entry $t(w)$ for each noun w drawing on WordNet[5] as follows. For each synset in WordNet, we select one member word to serve as the semantic feature representing that synset. The initial representation of $t(w)$ consists of the set of such features from each ancestor (hypernym) of the word's first sense in WordNet.[6] We use the same features as in previous work to initialize $t(w)$ for other parts of speech (Nematzadeh et al., 2011; Alishahi & Fazly, 2010).

To complete the representation of $t(w)$, we need a score for each feature which can be used in the probabilistic generation of a scene for an utterance containing w. We assume that general features such as ENTITY, that appear with many words, are less informative than specific features such as FOOD, that appear with fewer words. Hence, we aim for a score that gives a higher value to the more specific features, so that more informative features are generated more frequently.

We formulate such a score by forming semantic groups of words, and determining for each group the *strength* and *specificity* of each feature within that group; multiplying these components gives the desired assessment of the feature's informativeness to that group of words. [7]

First, we form noun groups by using the labels provided in WordNet that indicate the semantic category of the sense; e.g., the first sense of *apple* is in category *noun.food*. (For words other than nouns, we form single-member groups containing that word only.) Next, for each feature f in $t(w)$ for a word w in group g, the score is calculated by multiplying strength(f,g) and specificity(f):

$$\text{strength}(f,g) = \frac{\text{count}(f,g)}{\sum_{f' \in g} \text{count}(f',g)}$$

$$\text{specificity}(f) = \log \frac{|G|}{|g : f \in g|}$$

where $|G|$ is the total number of groups, and $|g : f \in g|$ is the number of groups that f appears in; strength(f,g) captures how important feature f is within group g (its relative frequency among features within g); specificity(f) reflects how specific a feature is to a group or small number of groups, with larger values indicating a more distinctive feature. For each word w, each feature f in $t(w)$ is associated with the score for f and g (where $w \in g$); the resulting scores are then

[4] It should be emphasized that the input-generation lexicon is not used for learning by the model; it is used only to create the input.

[5] http://wordnet.princeton.edu

[6] A native speaker of English annotated a sample of 500 nouns with their most relevant sense in our CDS corpus, revealing that the first WordNet sense was appropriate for 80% of the nouns. One regular exception was nouns with both 'plant' and 'food' senses, such as *broccoli*, which were predominantly referring to food. For these, we always use the 'food' sense.

[7] Our score is inspired by the tf-idf score in information retrieval.

re-scaled so that the maximum score is 1, to be appropriate for the probabilistic generation of the input scenes.

Experimental Results

In our previous work (Nematzadeh et al., 2011), we showed in computational simulations that LT learners not only learn fewer words than an ND learner, but that the LTs also have a less semantically-connected vocabulary, a result in line with the findings of Beckage et al. (2010). Here, using our extended model with its improved semantic representation, we first analyze the learned clusters of words for our two learners, to confirm that the semantic category knowledge of the LT learner is of substantially poorer quality. We also investigate the differential effects of the learned clusters for the two learners in subsequent word learning. It is known that word learning in children is boosted by their knowledge of word categories (Jones et al., 1991). Here, we interleave the two processes of semantic clustering and word learning in our model, and examine the patterns of word learning over time, for the two learners, with and without category knowledge. Our hypothesis is that the ND learner not only forms higher quality semantic clusters of words compared to the LT learner, but that its (more coherent) category knowledge contributes to improved word learning over time.

Analysis of the Learned Clusters

We examine the quality of the semantic clusters formed by each learner (ND and LT). We train the learners on 15K utterance–scene pairs, and perform a hierarchical clustering on the resulting learned meanings of all the observed nouns. To provide a realistic upperbound as a point of comparison for the two learners, we also cluster (using the same clustering algorithm and similarity measure) the true meanings of the nouns. These "TRUE" clusters indicate how well the nouns can be categorized by the clustering method on the basis of their true (in contrast to learned) meanings. In all cases, we set the number of clusters to 20, which is the approximate number of the actual WordNet categories for nouns.

To measure the overall goodness of each of the three sets of clusters (TRUE, ND, and LT), we compare the clustering to the actual WordNet category labels for the nouns, as follows. (The WordNet category labels reflect human judgments of semantic categories, since they are provided by manual annotation.) We first label each cluster c with the most frequent category assigned by WordNet to the words in that cluster, called label(c). We then measure *P(recision)*, *R(ecall)*, and their harmonic mean, *F(-score)*, for each cluster, and average these over all clusters in a set. Given a cluster c, P measures the fraction of nouns in c whose WordNet category matches the cluster label; R is the fraction of all nouns whose WordNet category is label(c) that are also in c. We report the average P, R, and F scores for the TRUE, LT, and ND clusters in Table 1.

As expected, the F score is the highest for the TRUE clusters, which result from the same clustering algorithm but applied to noise-free semantic representations. In comparison, the ND learner has somewhat lower F scores, as well as P and R scores, compared to the TRUE clusters. By contrast, the LT clusters have a very low F score. These results confirm that, in contrast to the ND learner, the LT learner is unable to use its learned knowledge of word meanings to form reasonable categories of words, confirming that nouns in the vocabulary of the LT learner have less semantic coherence than those of our ND learner. Moreover, the unusual nature of the clusters formed by the LT learner (in contrast with ND) is further confirmed by its very high P and very low R scores compared to the TRUE clusters. Detailed examination of the clusters reveals that LT has learned a large number of small clusters (leading to high precision), but also a few large semantically-incoherent clusters (leading to very low recall).

TRUE			ND			LT		
P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
.77	.71	**.66**	.79	.53	**.51**	.88	.19	**.24**

Table 1: Average P, R, and F scores (shown in boldface), for the TRUE, LT and ND clusters after processing 15K input pairs.

Incorporating Categories in Word Learning

Here we investigate the role of category formation in a naturalistic word learning setting. Specifically, we interleave the two processes by allowing the model to use its semantic clusters in word learning. To simulate the simultaneous learning of categories and word meanings, the model builds clusters from its learned noun meanings after processing every 1000 input utterance–scene pairs. It then uses these clusters when processing the next 1000 pairs (at which point a new set of clusters is learned). After the first 1000 input pairs, the model calculates the alignment probabilities using both word-based and category-based knowledge, as in Eqn. (4).

For each noun in an utterance, if it has been observed prior to the last clustering point, the model uses the cluster containing the noun to calculate the category-based alignment. But a novel (previously unobserved) noun has not yet been assigned to a cluster. However, it is recognized that children can use contextual linguistic cues to infer the general semantic properties of a verbal argument (Nation et al., 2003). For example, a child/learner knowing the verb *eat* might be able to infer that the novel word *dax* in "she is eating a dax" is likely referring to some 'edible thing'. We assume here that a learner can use the context of a novel noun to identify its general semantic category. In our model, we simulate this inference process by giving the model access to the WordNet category label of the novel word. Recall that each noun sense in WordNet is assigned a category label that provides information about its general semantics. The model can then choose a learned cluster for the novel noun by identifying the cluster whose assigned label matches the WordNet category of the noun. If more than one cluster has the same label as the category of the novel word, the cluster with the highest precision is selected. If the learner does not have a matching cluster, no category information is used for the novel word.

We process 15K input pairs overall, and look at the aver-

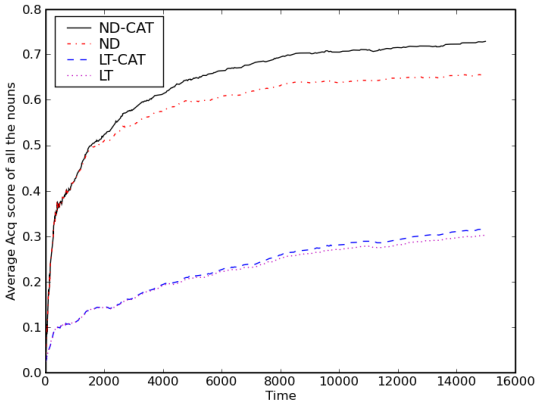

Figure 2: Change in the average Acq score of all nouns over time; ND-CAT and LT-CAT use category formulation during learning.

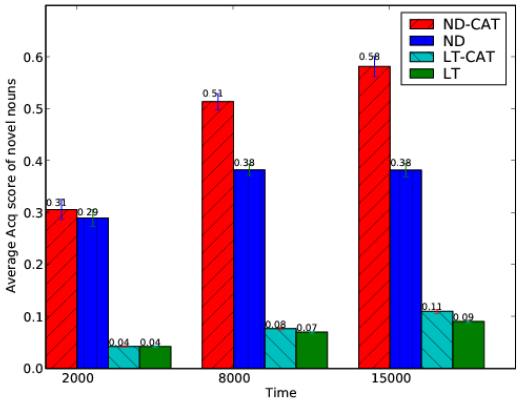

Figure 3: Changes in the novel word learning over time

age acquisition score (Acq, defined below) of nouns for each learner, with and without category knowledge, as a function of time (the number of input pairs processed); see Figure 2. The Acq score for a word w shows how similar its learned meaning $l(w)$ is to its true meaning $t(w)$:

$$\text{Acq}(w) \quad = \quad \text{sim}(l(w), t(w)) \tag{6}$$

where sim is the cosine similarity between the two vectors.

A comparison of the curves in Figure 2 reveals several interesting patterns. First, the use of category knowledge substantially improves the word learning performance of ND, whereas it has no effect at all on the (poorer) performance of the LT learner. These results further elaborate the findings of our analysis of the learned clusters: the clusters learned by the ND are a better match than those of the LT with the manually-annotated categories provided by WordNet; moreover, they are able to contribute helpful information to word learning, where the LT clusters are not.

Thus, the LT clusters are not only in principle of lesser quality, they are in practice less useful. Also, the positive effect of category knowledge for ND increases over time, suggesting that the quality of its clusters improves as the model is exposed to more input. This mutually reinforcing effect of semantic category formation with word learning underscores the importance of studying the interaction of the two.

Category Knowledge in Novel Word Learning

Results of the previous section suggest that the ability of a learner to form reliable categories of semantically-similar words may be closely tied to its word learning performance. In particular, we expect category knowledge to increase the likelihood of associating a word with its relevant semantic features when there is ambiguity and uncertainty in the cross-situational evidence. For example, when a child hears "The wug will drink the dax" while observing an unknown animal and a bowl of liquid in the scene, the child must rely on information sources other than the cross-situational evidence to infer the possible meanings of the two novel words. (That is, the child must infer that *wug* as a drinker is more likely to be the unknown animal.) We predict a substantial benefit of category knowledge when observing a word for the first time, since this is when there's the least cross-situational information available to a learner about the particular word and its features. Here we examine the effect of category knowledge on the learning of novel words over time, within the naturalistic setting of the utterance–scene pairs of our corpus, focusing on those inputs that include previously unseen words.

We train the model on 15K input pairs, but restrict evaluation to the learning of novel words. Specifically, we look at the difference in the Acq score of words at their first exposure, for the ND and LT learners, each with and without using category knowledge. To do this, we look at utterances containing at least two nouns, at least one of which is novel.[8] For each such input utterance, we record the resulting Acq score of all novel words in the utterance, and take their average. For each learner, we also examine the pattern of change in these average scores over time, as shown in Figure 3.

The results show that after 2K input utterances, there is no difference between using and not using categories for each of the learners (i.e., comparing ND-CAT and LT-CAT to ND and LT, respectively). This is because none of the learners has formed sufficiently good categories yet. After 8K utterances, ND-CAT performs much better than ND, showing the benefit of using category knowledge in learning novel words in an ambiguous setting. By contrast, for the LT learner, the Acq score of the novel nouns does not increase when using category information (LT-CAT) even with additional exposure to the input. Another interesting pattern is that for the ND learner, the average Acq score does not increase between 8K and 15K input utterances. However, when using categories (ND-CAT), this score increases over time. Although the ND model has learned additional words after 15K inputs, knowledge of more words alone does not result in improved learning of novel words. By contrast, the increasing semantic category knowledge in ND-CAT over time leads to greater improvements in learning the meaning of novel nouns.

[8] If the utterance only has 1 novel noun, the task is too easy because the features of nouns and other parts of speech do not overlap.

Conclusions

One possible explanation for the language deficiencies of late-talking children is inadequacies in their attentional and categorization abilities (Jones & Smith, 2005; Colunga & Sims, 2011). In this paper, we have investigated (through computational modeling) two interrelated issues: (1) how variations in the development of attentional abilities in normally-developing (ND) and late-talking (LT) children may interact with their categorization skills, and (2) how differences in semantic category formation could affect word learning. We have extended a model of word learning that incorporates an attention mechanism (Nematzadeh et al., 2011) to incrementally cluster words, and to use these semantic clusters in subsequent word learning.

Psycholinguistic findings have noted that the vocabulary of LTs shows both a lack of appropriate generalization (Jones & Smith, 2005; Colunga & Sims, 2011), and less semantic connectivity (Beckage et al., 2010; Sheng & McGregor, 2010). We find here that the clusters formed by our LT model indeed show more inconsistency and less coherence compared to our ND learner. In addition, unlike our LT learner, our ND model can use its learned knowledge of word meanings to form semantically-coherent and informative categories, which in turn contribute to an improvement in subsequent word learning. Moreover, the LT learner has particular difficulties in learning novel words, while the ND learner gets increasingly better over time when it draws on category knowledge. The inability of an LT learner to form reasonable semantic clusters limits its ability to generalize its knowledge of learned words to new words. This could be a substantial factor in the LT's delayed vocabulary acquisition.

The model presented here treats semantic category learning and word learning as two interacting but independent processes. In particular, the mechanism for incorporating category knowledge into word learning simply adds this knowledge as another factor in guiding the formation of word–feature associations. Our ongoing work is exploring a unified mechanism in which category knowledge is integrated into the attentional mechanism of the word learning model. Such an approach will enable us to further explore how specific correlations between semantic properties and abstract categories (such as shape–solid object) emerge from the input (for LTs and NDs), and how these affect subsequent word learning.

References

Alishahi, A., & Fazly, A. (2010). Integrating syntactic knowledge into a model of cross-situational word learning. In *Proc. of CogSci'10.*

Beckage, N., Smith, L. B., & Hills, T. (2010). Semantic network connectivity is related to vocabulary growth in children. In *Proc. of CogSci'10.*

Colunga, E., & Sims, C. (2011). Early talkers and late talkers know nouns that license different word learning biases. In *Proc. of CogSci'11.*

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., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *Int'l J. of Language and Communication Disorders*, *43*(4), 361–389.

Ellis Weismer, S., & Evans, J. L. (2002). The role of processing limitations in early identification of specific language impairment. *Topics in Language Disorders*, *22*(3), 15–29.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *J. of Memory and Language*, *53*, 258–276.

Jones, S., & Smith, L. B. (2005). Object name learning and object perception: a deficit in late talkers. *J. of Child Language*, *32*, 223–240.

Jones, S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*(3), 499–516.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed., Vol. 2: The Database). Erlbaum.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Hecke, A. V. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, *78*(3), 938–954.

Nation, K., Marshall, C. M., & Altmann, G. T. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *J. Experimental Child Psychology*, *86*, 314–329.

Nematzadeh, A., Fazly, A., & Stevenson, S. (2011). A computational study of late talknig in word-meaning acquisition. In *Proc. of CogSci'11.*

Paul, R., & Elwood, T. J. (1991). Maternal linguistic input to toddlers with slow expressive language development. *J. of Speech, Lang., & Hearing Research*, *34*, 982–988.

Rescorla, L., & Merrin, L. (1998). Communicative intent in late-talking toddlers. *Applied Psycholing.*, *19*, 398–414.

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

Sheng, L., & McGregor, K. K. (2010). Lexical–semantic organization in children with specific language impairment. *J. of Speech, Lang., & Hearing Research*, *53*, 146–159.

Stokes, S. F., & Klee, T. (2009). Factors that influence vocabulary development in two-year-old children. *J. of Child Psychology*, *50*(4), 498–505.

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.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb–argument structure: An alternative account. *J. of Child Language*, *28*, 127–152.