Title

What You Learn Is What You See: Using Eye Movements to Study Infant Cross-Situational Word Learning

Permalink

https://escholarship.org/uc/item/3rs333jd

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN

1069-7977

Authors

Yu, Chen
Smith, Linda B.

Publication Date

2008

What You Learn is What You See: Using Eye Movements to Study Infant Cross-Situational Word Learning

Chen Yu and Linda B. Smith (chenyu@indiana.edu)
Department of Psychological and Brain Sciences, and Cognitive Science Program, Indiana University
Bloomington, IN, 47405 USA

Abstract

Recent studies show both adults and young children possess powerful statistical learning capabilities to solve the word-to-world mapping problem. However, it is still unclear what are the underlying mechanisms supporting seemingly powerful statistical cross-situational learning. To answer this question, the paper uses an eye tracker to record moment-by-moment eye movement data of 14-month-old babies in statistical learning tasks. A simple associative statistical learning is applied to the fine-grained eye movement data. The results are compared with empirical results from those young learners. A strong correlation between these two shows that a simple associative learning mechanism can account for both behavioural data as a group and individual differences, suggesting that the associative learning mechanism with selective attention can provide a cognitively plausible model of statistical learning. The work represents the first steps to use eye movement data to infer underlying learning processes in statistical learning.

Keywords: word learning, language development, eye tracking, computational modeling

Introduction

There is growing interest in the idea of language learning as a form of data mining. Structure that is not obvious in individual experiences or small bits of data is derivable from statistical analyses of large data sets (Landauer & Dumais, 1997; Li, Burgess & Lund, 2000; Steyvers & Tenenbaum, 2005; see a review by Chater & Manning, 2006). These techniques have been shown to be powerful in capturing syntactic categories (Mintz, Newport, & Bever, 2002; Monaghan, Chater, & Christiansen, 2005), syntactic structures (Solan, Horn, Ruppin, & Edelman, 2005) and word boundaries (Christiansen, Allen, & Seidenberg, 1998). Also growing are suggestions (as well as relevant evidence) that infants and young children are powerful statistical learners who make what seem to be sophisticated statistical inferences from even limited data (Newport & Aslin, 2004; Tennebaum & Xu, 2000; etc).

What is not so clear, however, is the nature of underlying statistical learning mechanisms. The working assumption seems to be that learners more or less passively accumulate data and then apply special statistical computations to that data. In this paper, we consider an alternative, that at least one form of statistical learning shown by infants, is the product of moment-by-moment attention, itself inherently selective, dynamic, and via simple associative mechanisms, dependent on and indicative of learning. We make this case in the context of infants' cross-situational learning of names and referents; the approach is to use eye-tracking measures of attention during individually ambiguous training trials and a simple associative model to predict individual differences in learning.

Cross-situational word learning has been proposed as a solution to the uncertainty inherent in trying to learn words from their co-occurrences with scenes (Siskind, 1996; Yu & Smith; 2007). Scenes typically contain many possible referents, with speakers talking about and shifting their attention rapidly among the potential referents. This uncertainty is still considerable even if one assumes a learner biased to link names to whole objects (e.g., Markman, 1990). For example, as illustrated in Figure 1, a young learner may hear the words "bat" and "ball" in the ambiguous context of seeing both a BAT and BALL without any information as to which word refers to which scene element. However, although the learner may have no way of knowing from any such *single* learning situation which word goes with which referent, the learner could nonetheless determine the right mappings *if* the learner kept track of co-occurrences and non-occurrences *across situations*, and evaluated the cross-situational evidence for word-referent pairings in the proper way. Using the example in Figure 1, if the learner viewed a second scene while hearing the words "ball" and "dog" and if the learner could remember and combine the conditional probabilities of co-occurrences from across the two situations, the learner could correctly infer that "ball" maps to BALL.

In a recent study, Smith and Yu (2008) showed that 12- and 14-month old babies do this. They presented the infants with learning trials on which there were always two seen objects and two heard names but no information as to which name went with which object. From such individually ambiguous learning trials, the infants learned the mappings of 6 names to 6 objects and did so in a learning experience that lasted in total less than 4 minutes. The cross-trial word-referent statistics were the only information available to disambiguate those word-referent pairings. Thus the infants

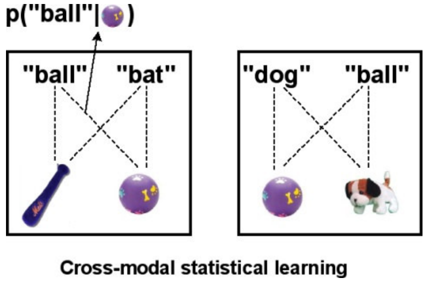

Figure 1: The conditional association probabilities between words and referents can be calculated across trials.

must have combined the information across trials. The present question is the nature of the processes that underlie this learning.

One possible learning process is Hebbian-like associative learning, a form of learning known to be fundamental to many perceptual and cognitive capabilities. In the present case, the learner could simply store all associations between words and references. For example, with respect to Figure 1, if the system stored only associations between words and whole objects, there would be four associations formed on trial one (bat to BAT, bat to BALL, ball to BAT, ball to BALL). On the second experience shown in the figure, one of these (ball to BALL) would be strengthened more than the others. Across trials, the relative strengths of associations between words and their potential referents would come to reflect the correct word referent mappings. Simple associative models such as this have been criticized on the grounds (see Keil, 1989) that there are just too many possible associations across situations to store and to keep track of.

This raises the key question for the present study, whether learners do not actually store all co-occurrences, but only some of them. Further, we ask whether infants' attention to and thus selective storage of word-referent pairs might be guided by their previous experience. And if this is so, could a simple associative model explain not only infants' success in learning in this task but also individual differences in that learning? The issue of individual differences is particularly critical if infants are not simply passive accumulators of data but instead select among the available data. If infants select some pairings over others to notice and store – and if these pairings guide later selections – then individual learners may distort the regularities in the input both in ways that enhance learning of the right word-referent pairs and in ways that hinder it.

To answer this question in computational modeling, a model needs to be fed with the same input that individual learners receive – that is, the information that individual learners attend to in cross-situational learning with multiple words and multiple referents. Our proposed solution is to continuously track eye-gaze direction throughout learning. The assumption here is when a learner associates a word with a referent among other simultaneously presented referents, the learner is likely to look at that referent (this is what it means to register the association). In this way, different learners may attend to different referents in a visual scene when hearing the same word, which leads to different learning results. Recent psycholinguistic studies already suggest that speech and eye movement are closely linked in both language comprehension and production (e.g. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

In brief, we apply this eye-tracking paradigm in language learning and use eye movement, and the synchrony of those movements with respect to the heard object names as a measure of moment-by-moment learning and as a clue to the internal state of the learner.

Method

Participants

The final sample consisted of 12 14-month-olds (7 boys, 5 girls), with a mean age of 14.3 (SD = 0.6) months. An additional 12 infants were tested but not included in the sample due to fussiness (n = 2), persistent inattention to the display n=2), and mostly occasional excessive movement that prohibited the complete collection of continuous eye movement data with the eye tracker (n = 8).

Stimuli

The 6 ''words'' *bosa*, *gasser*, *manu*, *colat*, *kaki* and *regli,* were designed to follow the phonotactic probabilities of American English and were recorded by a female speaker in isolation. They were presented to infants over loudspeakers. The 6 ''objects'' were drawings of novel shapes, each was a unique bright color. There were 30 training slides. Each slide simultaneously presented two objects on the screen for 4 s; the onset of the slide was followed 500 ms later by the two words – each said once with a 500 ms pause between. Across trials, the temporal order of the words and spatial order of the objects were varied such that there was no relation between temporal order of the words and the spatial position of the referents. Each correct word-object pair occurred 10 times. The two words and two objects appearing together on a slide (and creating the within trial ambiguities and possible spurious correlations) were randomly determined such that each object and each word co-occurred with every other word and every other object at least once across the 30 training trials. The first four training trials each began with the centered presentation of a Sesame Street character (3 s) to orient attention to the screen. After these first four trials, this attention grabbing slide was interspersed every 2–4 trials to maintain attention. The entire training – an effort to teach six word-referent pairs – lasted less than 4 min (30 training slides and 19 interspersed Sesame Street character slides). There were 12 test trials each lasting 8 seconds. Each test trial presented one word, repeated 4 times with 2 objects – the target and a distracter – in view. The distracter was drawn from the training set. Each of the 6 words was tested twice. The distracter for each trial was randomly determined such that each object occurred twice as a distracter over the 12 test trials. This duration and structure of training and test trials was the same as in Smith and Yu (2008).

Apparatus

A learner's eye gaze was measured by a Tobii 1750 eye tracker with an infant add-on (www.tobii.se). The principle of this corneal reflection tracking technique is that an infrared light source is directed at the eye and the reflection of the light on the corneal relative to the center of the pupil is measured and used to estimate where the gaze is fixated. The eye tracking system recorded gaze data at 50Hz (accuracy = 0.5°, and spatial resolution = 0.25°) as a learner watches an integrated 17 inch monitor with a resolution of 1280 x 1024 pixels.

Procedure

Infants sat in a parent's lap 60 cm from the 17'' computer monitor used to present the stimuli. Before the experiment a calibration procedure was carried out. In preparation for the calibration the experimenter adjusted the eye tracker to make sure that the reflections of both eyes were centered in the camera's field of view. We used a procedure including nine calibration points. The total duration of calibration procedure was about 3 minutes before the training trials start. Parents were instructed to look to the middle of the screen and not to interact with the child during the experiment.

Data

The eye tracker outputs (x,y) coordinates on the computer display of the visual presentation at the sampling rate of 50Hz. There are in total 120 sec (4 sec/per trial x 30 trials) during training and 96 sec (8 sec/per trial x 12 trials) during testing. Therefore, there are 6,000 data points in training and 4,800 data points in testing, if the eye tracker works perfectly. In practice, the tracking system occasionally failed to detect the subject's eye gaze either because the subject's head and gaze moved outside of the tracking plane, or the eye tracker could not correctly infer the subject's eye movements for some other reasons. For the 12 infants with good tracking results, the average tracking success is 76% in training and 71% in testing. Thus, on average, we collected 4,560 data points in training and 3,408 data points in testing per subject, which are used in the following data analysis and modeling.

Behavioral Results

Infants were presented with 30 training trials (two words and two objects) and then 12 test trials in which one target word was played and two objects (the correct referent and the distractor) were displayed. Infants' preferential looking on such test trials is used as a common measure of language comprehension in that infants systematically look at the portion of the display that corresponds to what they are hearing and this was the behavioral measure of learning used by Smith & Yu (2008). Accordingly, the first question we addressed was whether this study replicated the previous result: did infants during the test trials look longer at the correct referent for the heard word than the distractor? The answer is "yes," t(11) = 3.28, p < .01; thus, this study replicates the earlier finding that very young word learners can learn word-referent pairings from individually ambiguous learning experiences by combining the information across those experiences. The main purpose of this study, however, is to examine the relation between looking on the training trials and learning on the test trials. To this end, we first present a simple associative model that takes the micro-structure of the eye-tracking data during training as input and predicts individual performance on the test trials.

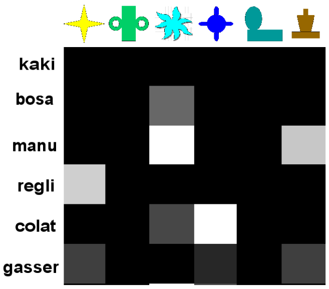

Figure 2: A 6 x 6 association matrix built based on the synchrony between a subject's eye movements and spoken words during training. Each cell represents the association probability of a word-object pair. The diagonal items are correct associations and other non-diagonal items are spurious correlation. Dark means low probabilities and white means high probabilities.

The Model

The model is conceptually very simple. An associative learning mechanism strengthens the link between a word and a referent if these two co-occur regularly across multiple trials and weakens the link if the word and the referent do not co-occur. In the current experiment, infants were exposed to 6 words and 6 pictures in total. Therefore, a 6 by 6 association matrix shown in Figure 2 is used in modeling as a representation of all the possible associations that a learner may keep track of. In such an association matrix, each cell corresponds to a particular word-referent association. The diagonal cells are the 6 correct pairings while non-diagonal cells represent spurious correlations due to the ambiguity inherent in training trials. The association probability of each cell is updated trial by trial in real-time learning. Given such an association matrix built through training, the learner can make a decision during testing simply by looking at the referent more strongly associated referent with a tested word. In this way, a successful learner would be one who built a matrix in which most diagonal items were assigned with higher probabilities than those in non-diagonal cells. In contrast, an unsuccessful learner would be one who accumulated strong but wrong associations between words and referents, those not on the diagonal. Thus the critical issue for learning is the specific associations that are accumulated over trials. A nonselective (ideal) learner would just keep track of everything. However, the more psychologically correct model may be one based on more selective attention and on attention that reflects current knowledge. Indeed, the very method of preferential looking to sights that *correspond* to heard words *assumes* that attention is so guided by knowledge.

To examine these possibilities, we suggest that the association matrix can be accumulated trial by trial as follows:

$$p_{ij}(t) = \frac{t-1}{t} p_{ij}(t-1) + \frac{1}{t} \frac{\lambda(t)\eta_{ij}(t)}{\sum_j \lambda(t)\eta_{ij}(t)}$$

t is the trial number, and $p_{ij}(t)$ refers to the association probability of the object i and the word j at the *t*th trial. Thus, $p_{ij}(t)$ corresponds to one cell in the association matrix which is composed of two weighted parts. The first part $p_{ij}(t-1)$ reflects the accumulated association probability so far until the (t-1)th trial that is carried over to the current trial. The second part (with two variables $\eta_{ij}(t)$ and $\lambda(t)$) updates the previous association probability based on a learner's eye movement in the current trial. More specially, we suggest that the dynamics of a learner's eye movements may reflect in two ways the learner's internal state during the current trial. First, rapid shifts of visual attention between possible objects after hearing a word may reflect the learner's uncertainty (that is, the lack of one stronger and one weaker association). In brief, we expect that the learner is more likely to consistently fixate on the corresponding referent to the degree that it is strongly associated with the target word; this is, again, the very basis of using preferential looking to measure word knowledge. This principle is encoded by $\lambda(t)$ that measures the overall degree of uncertainty in the tth learning trial from individual learners' perspective. Second, the multimodal synchrony between eye movements and spoken words may indicate the strength of the registration of a certain word-referent pairing, and the duration of such synchronized behaviours may indicate how strong that word-referent association is in the learner's association matrix. This observation is captured by $\eta_{ij}(t)$ that measures the possible association between a word i and an object j at the current trial based on eye movements. In the following, we explain exactly how to estimate $\lambda(t)$ and $\eta_{ij}(t)$.

We first computed eye fixations from raw eye movement data and converted the continuous gaze data stream into a set of eye fixations marked by the onset and ending timestamps of each fixation. Next, we segmented the whole set of eye fixations into individual trials by aligning eye fixations with the timing of training trials. Within each learning trial, there are multiple eye fixations on the two objects on the computer screen that occur as the two words are sequentially presented. Assume that there are L fixations $\{f_1, f_2, f_3, \ldots, f_L\}$ in the *t*th learning trial. For a fixation f_m, $v(f_m)$ is the object that was fixated on, $w(f_m)$ is the spoken word that the subject heard right before or during the fixation, and $T(f_m)$ is the fixation time. As shown in Figure 3, all of the eye fixations generated between the 200 ms after the onset of a spoken word and the onset of the next spoken word (or the end of the current trial) are assigned to be associated with that spoken word.

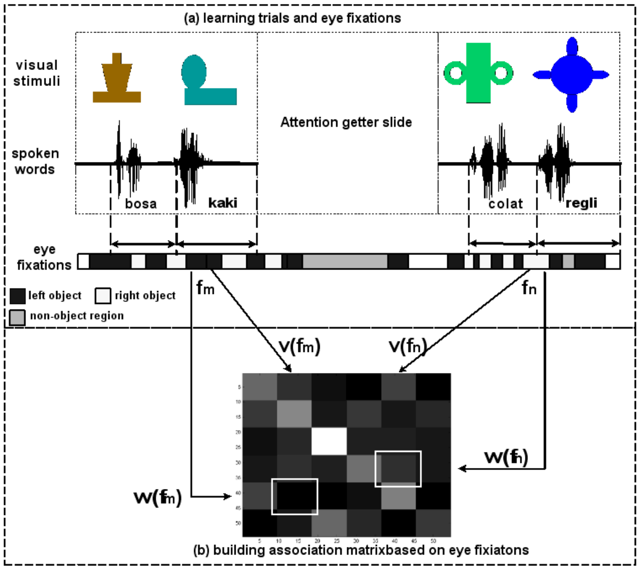

Figure 3: we measure where the learner is fixating on after hearing a spoken word. For example, after hearing the word "bosa", there are 4 eye fixations on both left and right objects. Those fixations (and corresponding fixed objects) are associated with the word "bosa". The strength of the association between an object (left or right) and the word "bosa" is determined by the overall duration of fixations on that particular object.

$\lambda(t)$, as an indicator of the learner's uncertainty in the current trial, can be encoded as how frequently the learner moves his eyes between those objects after hearing a word. Therefore, we use the entropy of a sequence of eye fixations within the trial as a metric to characterize this factor:

$$\lambda(t) = \sum_{m=1}^{L} \frac{T(f_m)}{\sum T(f_m)} log \frac{1}{T(f_m)/\sum T(f_m)}$$

where L is the total number of eye fixations within the *t*th trial.

Moreover, the second variable $\eta_{ij}(t)$ measures the possible association between a word and an object, which is composed of two parts. The first part estimates the probability of associating an object to a particular word based on the amount of time of looking at that object (compared with other objects) after hearing that word. Given multiple candidate objects, how likely is a heard word associated with each object. The second part estimates the probability based on comparing the looking time to the same object cross several spoken words. Given multiple candidate words, how likely is an object associated with each word. Formally, $\eta_{ij}(t)$ can be represented as follows:

$$\eta_{ij}(t) = \frac{\sum_{m=1}^{L} \delta(i, v(f_m)) T(f_m) \delta(j, w(f_m))}{\sum_{m=1}^{L} \delta(i, v(f_m)) T(f_m)} + \frac{\sum_{m=1}^{L} \delta(i, v(f_m)) T(f_m) \delta(j, w(f_m))}{\sum_{m=1}^{L} \delta(j, w(f_m)) T(f_m)}$$

where δ is the Kronecker delta function, equal to one when both of its arguments are the same and equal to zero otherwise. Thus, the denominator of the two parts is the same that accumulates the amount of fixations (T(fm), etc.) on a certain object j (v(fm) == i) after hearing a certain word j (w(fm) == j). The numerator in each part just normalizes the above denominator either cross all the words or cross all the objects respectively. Thus, a learner's visual attention in statistical learning is directly encoded in the association matrix the model built. Since individual infants generated different eye fixation sequences, the model builds different association matrices accordingly based on different inputs.

Results

Figure 2 shows an example of an association matrix built based on a learner's eye movements. In this example, some strong associations are correct (e.g., the word *manu* with the object manu) and others are not (e.g., the word *colat* with the object regli). Two measurements are used to evaluate whether the associative model based on eye movements can predict individual differences in statistical learning. First, we correlate the prediction of the number of learned words from the model with the number of learned words for each individual learner. There is a strong correlation between these two (p=0.71). Second, we also found the correlation between the proportion of diagonal cells (the strength of correct word-referent associations) in an association matrix with the proportion of looking time (the degree of the preference to look) at the correct referents during testing (p=0.65). Figure 4 shows the correlation between the model's prediction and the results from the empirical study. The four example association matrices built based on young learners' eye movements are quite different. Most diagonal items (correct word-referent associations) in the top association matrix are highlighted while the association probabilities between words and referents are more distributed in the bottom matrix. Critically, those matrices are built based on the same associative learning mechanism but with different eye movement data generated by subjects in real-time training. Thus, individual differences in this statistical learning task may be due to what infants attend to moment by moment while they all apply the same learning strategy.

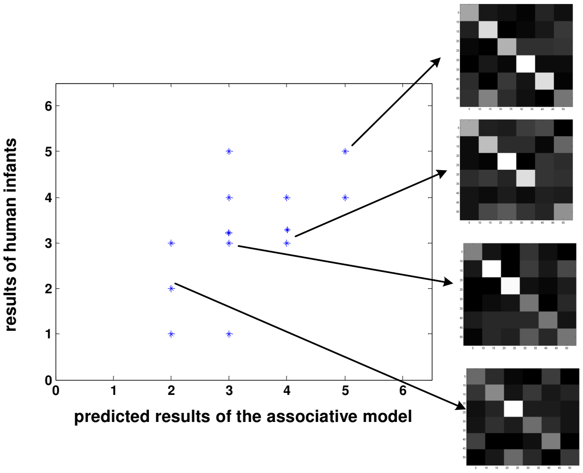

Figure 4: the comparison of predicted results from the associative model and the actual results of human learners indicates a strong correlation between these two.

General Discussion

For any learning mechanism to acquire knowledge from multiples instances separated in time, it needs to possess at least the following two components.

The first is **information selection.** The mechanism needs to accumulate data over multiple learning experiences. One option is to store nonselectively. If a learner did this, even if the learner randomly sampled the available co-occurrences, they would in the long run converge on an accurate representation of the regularities in the world. However, human learners' attention is unlikely to be random and, as shown here in infants, is guided by what they already know. This means that statistical learning will dynamically build on itself with each co-occurrence attended to influencing the probability (if it should occur again) that it will be attended to again. This kind of a system may both protect old learning and smartly direct attention to nonspurious co-occurrences. This first step of statistical learning from ambiguous contexts can play an important role in the following learning by selecting the right information and filter irrelevant data. However, not attending to the right co-occurrences could – at least temporarily – distort learning, sending it down the wrong path.

The second component of statistical learning is the **learning mechanism** itself. What kind of "computation" is used to evaluate the accrued data? The learning mechanism – as demonstrated here – could be as simple as associative learning that memorizes and keeps track of word-referent co-occurrences, or it could be as complicated as Bayesian graphical models using probabilistic inferences (Tenebaum & Xu, 2000). Moreover, the representation of learning results can be as straightforward as an association matrix or as complicated as relational hierarchical structures.

The present results suggest that infant cross-trial learning of word-referent correspondences can be explained by a simple learning mechanism coupled to selective attention. This contrasts with the more common approach to statistical learning which assumes sophisticated and powerful learning algorithms operating on messy data and most often running in a batch mode (e.g. Tenenbaum, etc.). Although these two accounts may be formally treated as variants of the same learning framework (Yu, Smith, Klein, & Shiffrin, 2007), a closer look also reveals the differences between two. An associative learning mechanism with real-time attention treats the learning process as a dynamical system and focuses on how the learning system may actively select the input based on real-time feedbacks from the current learning states and by doing so remove a significant amount of uncertainty from the data to facilitate the following processing. In contrast, the batch mode learning most often assumes that the learners perceive unprocessed ambiguous data to start with and then rely on the powerful learning machinery to infer meaningful knowledge. The first approach offers a potentially deeper and useful understanding of how learning progresses, how to promote, and how and why some learners are more effective than others.

In the long run, we need models and theories of learning that explain both information selection (as it dynamically happens in real-time learning) and also the learning mechanism itself. The present work is built upon the recent work in statistical word learning (Smith & Yu, 2008). Nonetheless, we go beyond demonstrating behavioral results and instead provide new insights into the underlying learning mechanisms. We do this by studying infants' attention during the course of learning, attention that is itself guided by learning. To achieve this goal, the current work is motivated by and takes advantage of three recent advances in cognitive science and psychology: (1) developmental psychology: using eye tracking techniques to measure moment-by-moment eye movement data from infants (Aslin & McMurray, 2006); (2) psycholinguistics: measuring the synchrony between visual attention and speech -- what are visually attended and what are heard (Tenanhaus, et al., 1999); and (3) computer science: modeling the learning mechanisms using computational techniques (Yu, Ballard, & Aslin, 2005). The work represents the first attempts to use momentary eye movement data as input to a computational model and by doing so to understand word learning processes. Indeed, our present results already suggest two important aspects in cross-situational learning. First, the results show that a simple associative learning mechanism can indeed work effectively and efficiently if the learner selectively registers the right statistical information at every moment. Second, different results from different learners may simply due to the fact that they attend to and select different statistical information encoded in the same training trials. Both observations are critical to understanding statistical learning processes. Here we show that eye movements can be used a window to infer the statistical learner's internal state, which allows us to ask in the future work how selective attention works in real-time learning. More specially, a generative dynamic model of selection attention can be integrated with the associative learning model here to provide a more complete picture of the underlying mechanisms.

Acknowledgments: We thank Juliette McNamara, Amara Stuehling, and Char Wozniak for collection of the data. This research was supported by National Institutes of Health R01 HD056029 and National Science Foundation Grant BCS0544995.

References

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy*, 6(2), 155-163.

Chater, N. and Manning, C.D. (2006) Probabilistic models of language processing and acquisition. *Trends in Cognitive Science*, 10(7), 335-344.

Christiansen, M., Allen, J., & Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13, 221-268.

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by. one-year-olds leads to specific and abstract knowledge. *Cognition*, 70,. 109–135

Fiser, J., & Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499-504

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning Mechanism. *Cognition*, 83, B35-B42.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Landauer, T.K., & Dumais, ST. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-140.

Li, P., Burgess C., Lund, K. (2000). The acquisition of word meaning through global lexical co-occurrences. *In Proceeding of Thirtieth Stanford Child Language Research Forum*, pp. 167-178.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1), 57-77.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26, 393-424.

Monaghan, P., Chater, N. & Christiansen, M.H. (2005). The differential role of phonological and distributional cues in grammatical categorization. *Cognition*, 96, 143-182.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48,127-162.

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

Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39-61.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition,* 106(3), pp 1558-1568.

Solan Z., Horn H., Ruppin E., and Edelman S. (2002). Unsupervised learning of natural languages. *Proceedings of National Academic of Science*, 102,11629-11634.

Steyvers, M. and J. B. Tenenbaum (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, 29(1), 41-78.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.

Tenenbaum, J., & Xu, F. (2000). Word learning as Bayesian inference. In L. Gleitman & A. Joshi (Eds.), *Proceeding 22nd annual conference of cognitive science society* (p. 517-522). Mahwah.NJ: ErIBaum.

Yu, C., Ballard, D.H., & Aslin, R.N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science, 29* (6), 961–1005.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414-420.

Yu, C., Smith, L. B., Klein, K., Shiffrin, R.M. (2007). Hypothesis Testing and Associative Learning in Cross-Situational Word Learning: Are They One and the Same? In McNamara & Trafton (Eds.), *Proceeding 29nd annual conference of cognitive science society* (p 737-742). Mahwah.NJ: ErIBaum.